M.TECH (POWER ELECTRONICS & SYSTEMS)

VIVEK PATHANIA

2022

DEVELOPMENT OF SOLAR POWER MANAGEMENT SYSTEM USING DEEP REINFORCEMENT LEARNING

DISSERTATION/THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
POWER ELECTRONICS & SYSTEMS

Submitted by:

VIVEK PATHANIA

2K20/PES/24

Under the supervision of

Mrs. SHATAKSHI
(Assistant Professor, EED, DTU)

DEPARTMENT OF ELECTRICAL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

2022

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi- 110042

CANDIDATE'S DECLARATION

I, Vivek Pathania, 2K20/PES/24, student of M.Tech. Power Electronics and Systems, hereby declare that the project dissertation title "Development of Solar Power Management System Using Deep Reinforcement Learning" which is submitted by me to the department of Electrical Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **VIVEK PATHANIA**

Date

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi- 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Development of Solar Power Management System Using Deep Reinforcement Learning" which is submitted by Vivek Pathania, 2K20/PES/24, Department of Electrical Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carrier out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi **Mrs. SHATAKSHI**

Date: **SUPERVISOR**

ASSISTANT PROFESSOR

ABSTRACT

Energy consumption has been steadily growing and is expected to persist in the future. This increasing demand for energy encourages the fast growth of renewable energy sources for instance solar, wind, tidal, geothermal, and others in order to reduce fossil fuel usage and keep environment pollution free. Aside from wind power, solar energy is the most widely utilised energy source, having a large market share in the global energy business. In recent times with significant developments in the field of photovoltaic (PV) panels the costs for PV systems have also been reduced. With advancements in power electronics converters for better system performance, an efficient maximum power point tracking (MPPT) controller is required to increase system throughput. To ensure that the maximum power point (MPP) can always meet the objective under varying weather circumstances of solar radiation and temperature, the MPPT algorithm is used in conjunction with a DC/DC converter. Along with a bidirectional converter (BDC) coupled with the battery system for utilising the stored power in times of less or no solar irradiance. This solar converter typically consists of DC-DC boost converter as the MPP converter and a bidirectional buck-boost converter as BDC. In modern times with the availability of faster computing devices and developments in artificial intelligence (AI), new and more adaptable MPPT controllers can be designed around AI. One such method is deep reinforcement learning which uses deep neural networks to model out the state-action in any given environment. Using

a deep deterministic policy gradient algorithm in MPPT ensures a continuous action space available for controlling the MPP converter and BDC. The MPP converter is responsible for achieving the maximum operation condition for solar modules whereas, the BDC is accountable for maintaining a steady DC bus voltage through charging and discharging of battery system. Loads such as DC loads and AC loads with a suitable inverter can be coupled to this DC bus to be utilised in household applications or electric vehicle charging stations.

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi- 110042

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to Mrs. Shatakshi for her continuous guidance and mentorship that she provided me during the project. She showed me the path to achieve my targets by explaining all the tasks to be done and explained to me the importance of this project as well as its industrial relevance. She was always ready to help me and clear my doubts regarding any hurdles in this project. Without her constant support and motivation, this project would not have been successful.

Place: Delhi **VIVEK PATHANIA**

Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PV	Photovoltaic
MPPT	Maximum power point tracking
MPP	Maximum power point
BDC	Bidirectional converter
AI	Artificial intelligence
DRL	Deep reinforcement learning
FFT	Fast Fourier transform
ReLU	Rectified linear unit
P&O	Perturb & observe
IC	Incremental conductance
FLC	Fuzzy logic control
ANN	Artificial neural network
DDPG	Deep deterministic policy gradient
PSC	Partial shading conditions
GA	Genetic algorithm
DQN	Deep Q network
SPMS	Solar power management system
Si	Silicon
GaAs	Gallium Arsenide
PWM	Pulse width modulation
S-PWM	Sinusoidal pulse width modulation
NN	Neural network
P-V	Power-voltage

LIST OF SYMBOLS

I_o	Reverse saturation current
V_D	Diode voltage drop
I_{pv}	PV cell output current
K_1	Current temperate coefficient
T_c	PV cell operating temperature
V_{in}	Input voltage
V_o	Output voltage
V_L	Inductor voltage
V_C	Capacitor voltage
d	PWM duty ratio
K_P	Controller proportional constant
K_I	Controller integral constant
I_{ph}	Current generated by photovoltaic cell
I_{sh}	Parallel resistance current in equivalent PV cell
I_L	Inductor current
T	Pulse time period
f	Switching frequency
V_{bat}	Battery terminal voltage
t_{max}	Maximum operation time for a battery
n	Transformer ratio
f_o	Cut-off frequency
A,a	Action
s	State
r	Reward
p(A)	Probability of Action
s_t	State, action pairs for an environment
N	Noise added while training of DDPG agent
I_D	Diode current

CHAPTER 1

INTRODUCTION

Today as the energy demand increases so does the pollution that comes from energy generation using traditional sources so, an alternative to traditional source has to be searched for in terms of solar, wind, etc. In tropical countries where there is an abundance of solar irradiance, solar energy can be the alternative to traditional energy sources. To utilize this solar energy photovoltaic (PV) systems are required for commuting the solar energy into electrical energy. The PV systems are outfitted with a maximum power point tracking (MPPT) system for tracking the maximum operating conditions for maximum power generation. The power generated by a PV system depends upon the available solar irradiance which may lead to power fluctuations in low irradiation condition thus, to oppress these fluctuations the PV systems are coupled with wind energy systems, grid connected systems, diesel generator systems or battery systems.

Coupling of battery and PV system helps in reducing the power fluctuation that arise due to the variations in solar irradiation, in low irradiation conditions the battery systems can provide the excess power demanded by the loads. The battery systems can be coupled at the DC side where the PV system and battery systems deliver power to the loads through a common DC bus or, coupling the battery systems to a AC-DC converter at the AC side where PV systems are used to operate the AC systems.

The MPPT system uses various MPPT algorithms that can be categorised into several groups based on sensor requirements, sturdiness, reaction speed, efficacy, and memory. Soft computing approaches such as fuzzy logic control (FLC), artificial neural network (ANN), and neuro-fuzzy control are another type of MPPT control. Methods built on soft computing methods and evolutionary algorithms can cope with the nonlinear issue quickly and find global solutions however, for low convergence randomness, it

usually necessitates the use of a costly microcontroller and greater understanding of a specific PV system.

When compared to meta-heuristic approaches, RL has greater convergence stability and takes less time to compute, making it a possible tool for addressing the MPPT control issue optimally. A Q-learning controller learns the ideal duty cycle for each control region based on environmental interactions. However, in complex systems with high number of state and action space the computational time may tend to increase thus affecting the convergence rate. To oppress this deep reinforcement learning (DRL) can be employed, it provides a large state space and continuous action space for complex systems. DRL replaces the Q-value table used in RL with a neural network for function approximation thus, eliminating the need for large Q-tables that require too much time in complex systems. In this a model-free DRL algorithm deep deterministic policy gradient (DDPG) is employed that provides a continuous action space rather than a lookup table-based approach in RL.

The DDPG based controller is responsible for the control of MPPT system and energy storage system. The MPPT system consists of a boost converter as the maximum power point (MPP) converter whereas, the energy storage system comprise of a bidirectional buck-boost converter as the bidirectional converter (BDC). The MPP converter and BDC are coupled together at the DC bus where loads can be attached such as DC loads or AC loads with the help of an inverter system. The MPP converter is responsible for operating the PV array at MPPT whereas, the BDC is responsible for keeping the DC bus voltage constant by utilising the battery.

CHAPTER 2

LITERATURE REVIEW

2.1. OBJECTIVE OF WORK

As solar power shows great scope in meeting the energy demands of present and future, more developments have conducted in this area. Such developments have led to decrease in costs of photovoltaic modules and solar power converter power electronics [1]. To enhance maximum power output is achieved by the solar power converter a maximum power point tracking (MPPT) algorithm is utilized. Various MPPT algorithms have been used throughout the years, and they may be categorised into several groups based on sensor requirements, robustness, reaction speed, efficacy, and memory [2]. The objective of solar PV systems is to converter maximum amount of solar energy into electrical energy with maximum tracking efficiency and speed.

2.2. SOLAR PV SYSTEMS

Solar PV systems can be coupled to various other systems such as wind, grid connected or diesel generation to reduce the fluctuations in solar power. To tackle the challenges associated with both PV and wind generators (WG), a hybrid system integrating the two is preferred. Indeed, combining wind and PV generator systems appears to overcome the disadvantages of their unpredictability and liability to weather and climate change [3]. To meet power needs in remote places, it is practical to design a system consisting of WG and PV panels [4]. Solar PV systems can also be couple with the grid to acts as distributed generation units. The term "distributed generation" refers to the supply to individual area from solar panels which are installed in each of these areas. These areas are also connected to the grid, allowing them to draw

power from the grid when demand exceeds PV output or feed power back into the system when PV output exceeds demand [5]. The grid connected inverter is at the base of any grid-connected power system [6]. The current control mode of a grid-connected inverter operating in grid is basically a voltage source input and current source output of the inverter [7]. A grid-connected PV system usually has several inverters, it's vital to look at the control methodology of a single inverter as well as the reactive power output limits [8].

2.3. SOLAR PV BATTERY SYSTEMS

All renewable energy systems have shortcomings, the most significant of which is their unpredictable output and forecasting. As a result of their complementing influence on each other, with the right control approach and hybrid renewable systems, it is possible to produce power with a high degree of assurance [9]. Using energy storage devices to compensate for a lack of energy is also a useful way to reduce the uncertainty of renewable energy sources [10]. There are three types of PV storage systems that are connected as common DC bus (PV-battery hybrid system), common AC bus, and AC/DC hybrid bus linked [11]. The common DC bus connected storage is more effective and simpler to control the power flow, making energy management more straightforward. A common AC bus connected storage system requires a simplified control system that can remould the shape of an existing PV generation system but with poorer efficiency [11-13]. While the AC/DC hybrid bus linked storage combines the benefits of both DC and AC bus connected storage with a more complex design [14-16].

2.4. MAXIMUM POWER POINT TRACKING ALGORITHMS

The noteworthy algorithms are perturbation and observation (P&O) [17] and incremental conductance (IC) [18]. Under homogeneous solar radiation conditions, traditional approaches may generally work well however, their significant disadvantage is that they get caught up in local maximum power point (MPP), which results in limited energy conversion under partial shade conditions (PSC) [19-20]. Furthermore, a small duty cycle step size results in prolonged tracking time, but a big duty cycle ratio makes PV power to fluctuate about the MPP [24-26]. To overcome this improved P&O (Improved P&O) algorithm is also developed which has a faster MPPT speed and efficiency but, requires a microcontroller of higher advancement [27]. Current issues in

power system control include generating unit scheduling and control in a stochastic environment, as well as regulatory approaches to handle the growing number of power conversion stages [28-29]. Traditional power scheduling approaches such as dynamic programming (DP) [30], linear programming (LP) [31], have been used. These techniques, however, have seem to face the dreaded dimensionality bottleneck and are unable to respond to the optimization environment's stochasticity, which includes volatile load profiles, grid tariffs, and renewable energy sources generation [32-34]. Apart from the traditional algorithms some other approaches, such as the genetic algorithm (GA) [35], cuckoo search [36], are based on evolution algorithms and fuzzy logic control are based on soft computing methods have also been developed. In case of cuckoo search higher convergence speed and more robustness with lesser variables is observed [37]. Power management problems have also been solved using global search approaches such as genetic algorithms (GA) and swarm intelligence (SI) [38]. However, because these approaches are sluggish, they cannot be used online. offline optimization does not need a reserved computer, online execution allows for more efficient use of computing resources [39-41]. There is no learning component in the methods listed above. As a result, every new load and generation profile necessitates optimization iterations, which is computationally expensive. Furthermore, to anticipate state variables, a separate forecasting algorithm is used. [42-44].

2.5. DEEP REINFORCEMENT LEARNING TECHNIQUES

Reinforcement learning (RL) algorithms are a exceptional choice since they can be learned for a generic load and generation profile offline and then use the learned values online to any load and generation profile. Whereas, other optimization approaches such as numerical methods and soft computing techniques, RL does not involve an exact model of the optimization environment to arrive to an optimal solution. [45-46]. Extensive research has been the point of focus in the field of reinforcement learning (RL) with various successful applications, owing to its high learning ability from environmental interactions and historical data, rather than the employment of complex mathematical models for development of control system [47]. RL in comparison to meta-heuristic methods shows significant reduction is computation time and higher convergence rate, the most common of all RL based methods is the Q-learning based method where the state and actions are mapped out by the environmental interactions

[48]. However, in case of complex systems with large number of possible states or actions the computational time is severely affected. The outcome of these problems gave rise to a new method in RL based learning called deep reinforcement learning (DRL) [49]. When it comes to complex state space scenarios DRL shows a better performance capability, as the complex state-action table used in RL is replaced by a neural network thus making it possible to have continuous state space [50]. The most predominant DRL algorithms in MPPT are deep Q network (DQN) and deep deterministic policy gradient (DDPG), the DQN provides a discrete action space where as the DDPG provides a continuous action space [51-53].

CHAPTER 3

MODELLING OF SYSTEMS

3.1. INTRODUCTION

The solar power management system (SPMS) can be divided into sub-systems based upon the action performed by the sub-system. These sub-systems then work as a whole to support the applications of the SPMS from conversion of solar energy to storage of electrical energy to consumption of solar energy. The solar power conversion takes place by the help of PV modules which convert the solar irradiation into solar power. With the help of a MPP converter, the PV modules are able to operate at maximum power point delivering maximum solar power at a particular irradiance level. The storage for solar power takes place in the energy storage system where a battery is used for energy storage along with a bidirectional converter that aids in power flow between the battery and DC bus in charging and discharging operations. The loads connected to the DC bus act as the utilizer of solar power for various applications at household level or electric vehicle charge stations.

The sub-systems of SPMS can be differentiated as solar power tracking system, energy storage system and applied loads. The solar power tracking system consists of PV module coupled to a maximum power point (MPP) whereas, the energy storage system is comprised of a bidirectional converter that is capable of operating the battery in charging and discharging modes. The applied loads can further be divided into DC loads and AC loads where an additional inverter in required for operating the AC loads. These sub-systems are attached to a common DC bus which acts a common point for power transfer between sub-systems. Figure 3.1, represents the power flow that occurs in the SPMS where

unidirectional power flows between the PV module and DC bus but, bidirectional power flow is achieved at the energy storage subsystem.

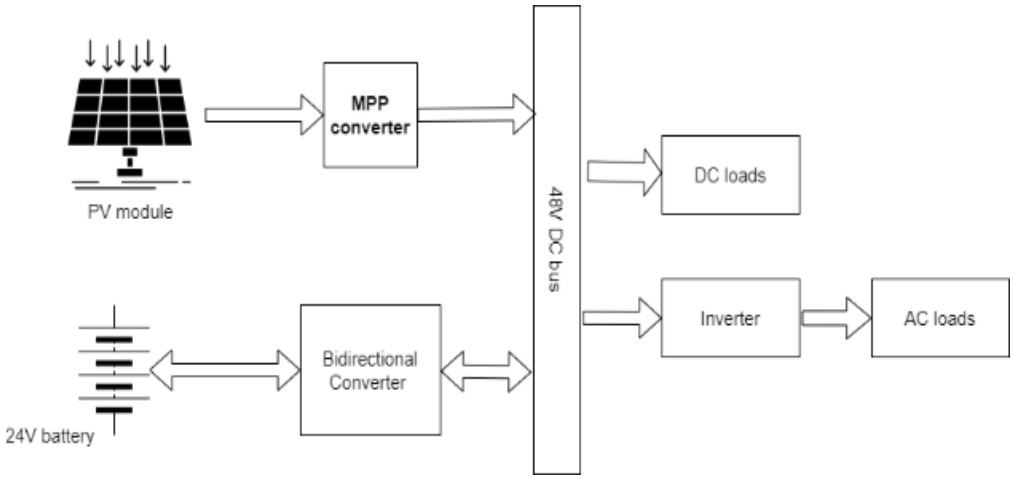


Fig 3.1: Power flow of PV system with Bidirectional converter

3.2 SOLAR POWER TRACKING SYSTEM

The solar power tracking system consists of PV modules that are a collection of PV cells and MPP converter, represented by figure 3.2. The MPP converter is responsible for tracking the maximum solar power that can be attained from the PV module at a particular irradiance level.

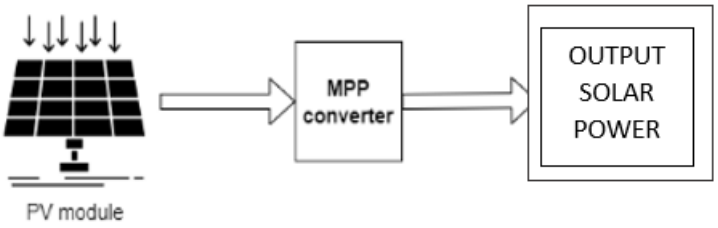


Fig 3.2: Block diagram of solar power tracking system

3.2.1. PV cells

The most popular solar cells are huge p–n junction diodes that generate DC power from light energy (photons). Similar to a diode which is an electrical device that permits unidirectional flow of current. The solar cell is made up of n- and p-layers that come together to form a junction. Compounding doped semiconductor materials like Si or GaAs forms the p–n junction.

In a solar cell, energy conversion takes place in two phases. First, an electron–hole pair is formed by the absorption of light of the suitable wavelength. Light absorption is indeed the elimination or absorption of photons caused by the stimulation of an electron from the valence band to the conduction band. The n-type material allows electrons to flow freely, whereas the p-type material allows holes to flow freely. The device's electrical structure separates the light-generated electron and hole, with electrons going to the anode and holes to the Metal contacts that forms the positive terminal on the front and rear of the cell gather electrical power. The rear contact is usually solid metal, whereas the front contact is a metal grid. The existence of electrons and holes generates net negative and positive charges, resulting in an electric field around the metallurgical junction. The depletion area is created when the electric field removes the electrons and holes.

The lighted features of the solar cell are coupled with mathematical calculations from the ideal diode equation to produce the ideal current-voltage characteristics of a p–n junction diode when illuminated by light. The ideal diode law is expressed as the following equation:

$$I_d = I_o(e^{\frac{qV_D}{KT}} - 1) \tag{3.1}$$

where q represents the electric charge constant, I_0 is the reverse saturation current, and V_D is the voltage across the diode terminals. A solar cell equivalent electrical circuit of a single-diode model based on Kirchhoff's law can be represented as shown in figure 3.3, the output current of an ideal cell is given by:

$$I_{pv} = I_{ph} - I_D - I_{sh} \tag{3.2}$$

I_{sh} is the parallel resistance current, given by

$$I_{sh} = \frac{V + I_{pv}R_s}{R_{sh}} \tag{3.3}$$

I_{ph} represents the light generated current proportional to intensity of the light. It is calculated by

$$I_{ph} = [I_{sc} + K_I(T_c - T_r)] \times \frac{G}{G_{STC}} \tag{3.4}$$

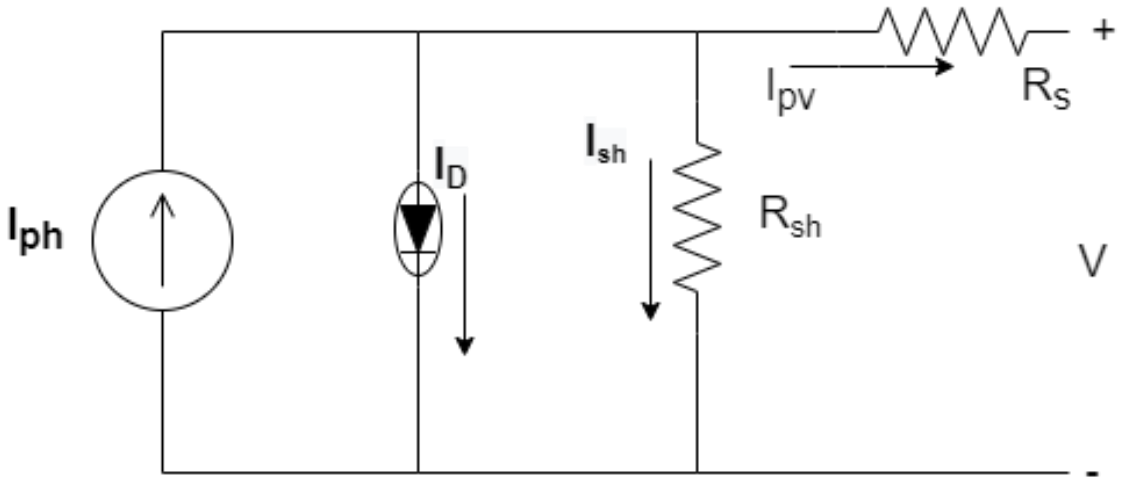


Fig 3.3: Solar cell Equivalent circuit

From equations 3.1, 3.2 & 3.3.

$$I_{pv} = I_{ph} - I_o(e^{\frac{q(V+IR_S)}{AKT_cN_S}} - 1) \quad (3.5)$$

where I_{sc} represents the short-circuit current, G denotes the relative irradiation, K_1 is the current temperature coefficient at short-circuit conditions, T_c accounts for the cell's operating temperature, and T_r represents the temperature reference.

3.2.2. PV Module

A PV module is a collection of cells that are linked electrically in series or parallel. The modules electrical properties are similar to those of individual cells, with power, current, and voltage varying depending on how many cells are linked in series or parallel. A collection of such modules forms a PV panel which can also be connected electrically as in series or parallel and collection of these panels creates a PV array. This can be analysed by the help of figure 3.4, where the array is the largest unit and cell is the smallest unit in a PV array.

Each module has two output terminals that capture and transmit the generated current to the solar power management systems. The ratio between the electrical power going out of the terminals and the power of the radiation from the sun impacting the module's surface is used to calculate the efficiency of a photovoltaic module. The standard amount for solar radiation is 1,000 W/m^2. When 1,000 watts of sunlight strike a square metre, the efficiency is the percentage of that energy and the energy that is transformed into useful power.

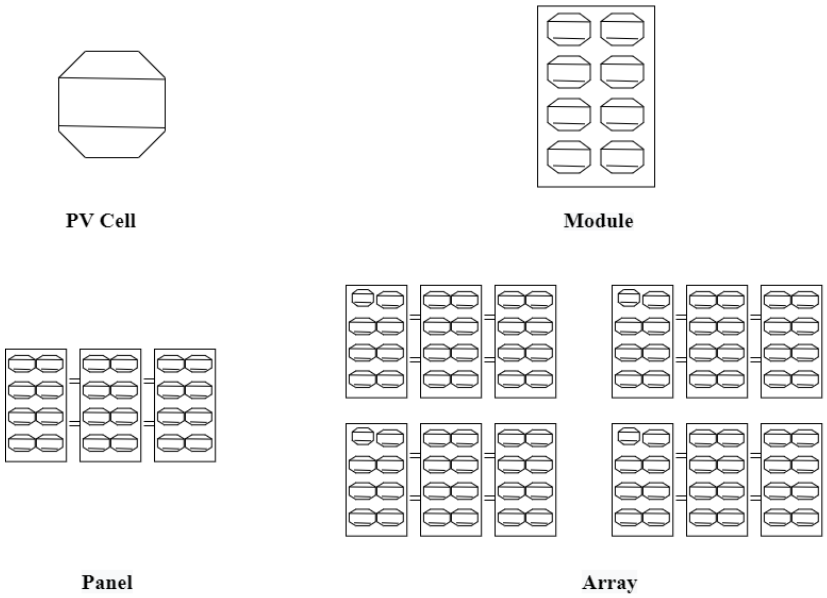


Fig 3.4: Difference between PV cell, module, panel & array

3.2.3. Maximum Power Point Converter

The MPPT method relies heavily on DC–DC converters. When a PV array's output terminals are connected to a DC–DC converter, the PV array voltage is regulated by adjusting the duty cycle (d), which is a pulse width modulated (PWM) signal that the MPPT controller uses to manage the voltage at which maximum power point cab be achieved. A DC-DC boost converter is used as a MPP converter which works as a step-up converter where the output voltage (V_o) is maintained greater than the input voltage (V_{in}) by operating an induction to store energy momentarily.

For a DC–DC boost converter, the V_0 is obtained by referring to the figure 3.5(a), the switch S1 is kept in a closed position thus creating a short across the inductor and battery. Here the current flows through the inductor (L) and a voltage are observed across this L as V_L. In this case the load is supplied through the capacitor C with a voltage across it as V_c.

$$V_{in} = V_L \quad , \; i_c = I_L \tag{3.6}$$

V_L, i_c can be written in terms of change in inductor current (i_L) and capacitor voltage

$$V_{in} = L\frac{di_L}{dt} \tag{3.7}$$

$$C\frac{dV_c}{dt} = \frac{V_o}{R} \tag{3.8}$$

Writing $\frac{di}{dt} = \frac{\Delta i_l}{DT}$ in equation 3.7,

$$L\Delta i_L = V_{in}DT \tag{3.9}$$

where D is the duty ratio and T represents the time period.

From figure 3.5(b) when the switch is on open position and diode D1 is closed, here the current flows from battery to the load through the inductor.

$$V_{in} = -V_L + V_o \tag{3.10}$$

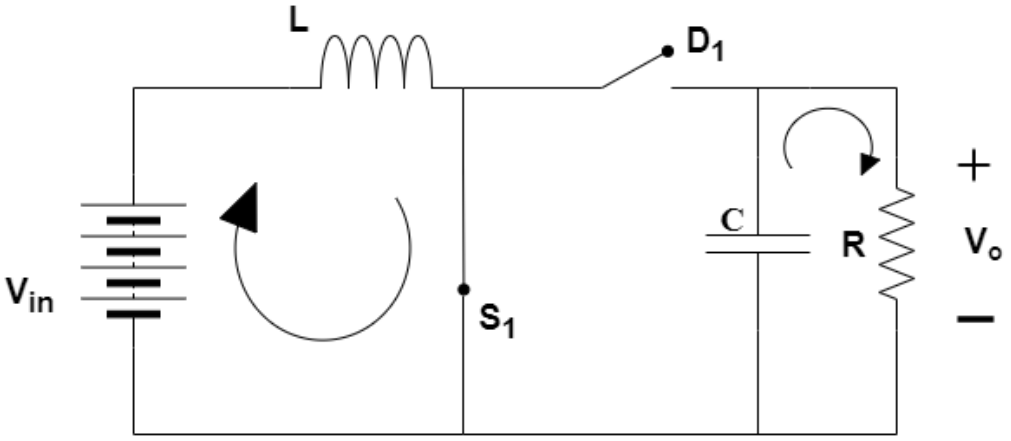
$$V_{in} = -L\frac{di_L}{dt} + V_o \tag{3.11}$$

From equation 3.9 and replacing dt = (1-D) T,

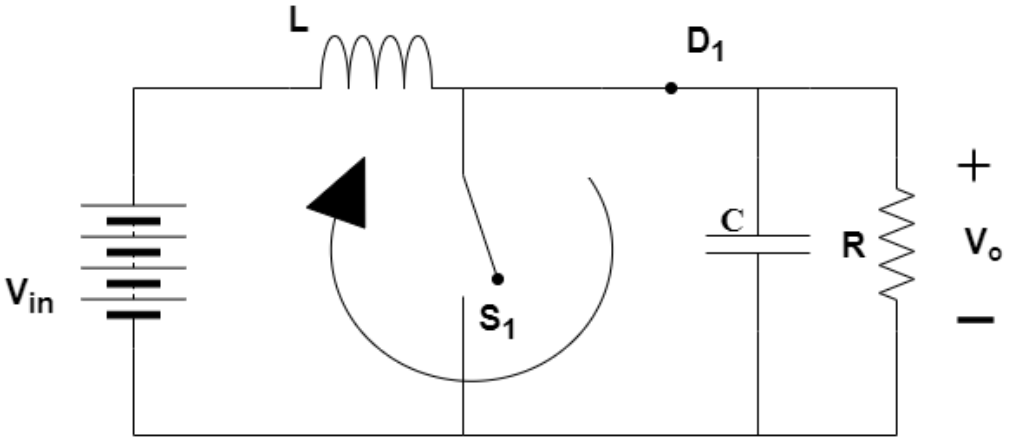
$$V_{in} = -\frac{V_{in}D}{1-D} + V_o \tag{3.12}$$

Rearranging equation 3.12,

$$V_o = \frac{V_{in}}{1-d} \tag{3.13}$$



(a)



(b)

Fig 3.5: operation of boost converter, (a) switch S1 in close & D1 in open, (b) switch S1 in open & D1 in close

According to figure 3.6, the model of MPP converter can be used to analyse the working of the MPP converter, when switch S1 is closed the current flows from the source to the inductor helping in storing the magnetic energy within the inductor (L_{boost}). When S1 is opened the stored energy flows from the source along with the magnetic energy stored in

the inductor towards the load through diode (D). The inductor has to be selected accordingly to help in the process of boosting the input voltage, the inductor is made to function in continuous conduction mode to draw the current from PV array continuously. So, for selecting the appropriate value for the inductor following formula is used,

$$L_{boost} > \frac{d(1-d)^2 R_{load}}{2f} \quad (3.14)$$

Here R_{load} represents the equivalent load, f is the switching frequency.

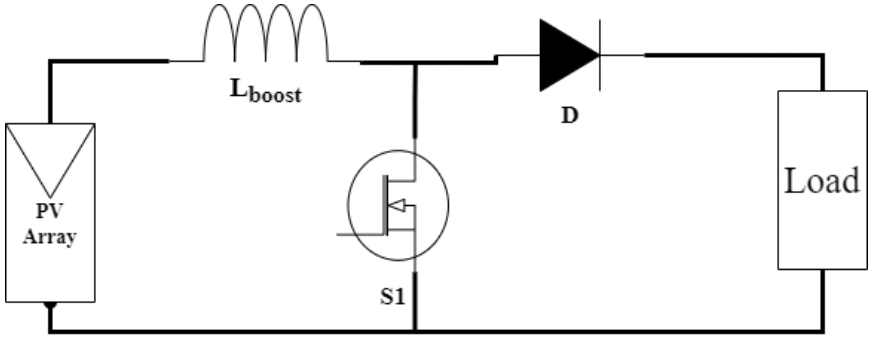


Fig 3.6: Model of MPP converter

3.3. ENERGY STORAGE SYSTEM

The energy storage system constitutes of battery along with a BDC converter shown by figure 3.7 that is responsible for charging and discharging of the battery system. It supports a bidirectional power flow where the power obtained from the PV array can be stored hence providing battery charge operation or delivering power to the load hence providing battery discharge operation.

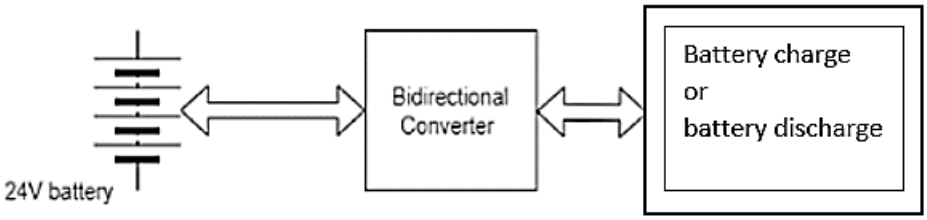


Fig 3.7: Block diagram of energy storage system

3.3.1. Battery

Batteries are used in PV energy storage systems to store and release electrical energy in times of low power demand and high power demands respectively. While considering a battery to be used as an energy storage device in the PV system, the rating

of this battery has to be kept in mind which is represented in kilowatts (kW). During high load conditions, the battery has to meet the load demands, in this scenario the rating of battery should be enough to support the load demand for a specific duration of time. Batteries of different power rating can be categorised as continuous power rating or instantaneous power rating. A continuous power rating battery is capable of supply a high load at constant power continuously whereas, an instantaneous battery is capable of suppling high power momentarily for a short duration and then operating at low power continuously. Other important factor when selecting a battery is the battery capacity, which is the ability of the battery to store and supply the power and is expressed as ampere-hour (Ah). Battery capacity tells how long the battery power will last, if high power is demanded from the battery, then this battery will last for a less amount of time whereas, low demand ensure longer battery duration. Battery capacity required for the energy storage system can be calculated as:

$$capacity\ (Ah) = \frac{P_{max}\, t_{max}}{V_{bat}} \tag{3.15}$$

P_{max} represents the maximum load in kW that the DC bus is rated for, t_{max} represents the maximum duration in hours the load can be operated for & V_{bat} represents the battery terminal voltage.

The last and most important factor while selecting a battery is the battery chemistry which generally means the composition of the battery, a lithium-ion battery is made using lithium metal whereas a lead-acid battery is made from lead metal. These are the two most commonly used battery chemistries in solar power applications with each have their advantages and disadvantages. Lithium-ion battery has a higher capacity, higher efficiency, high depth of discharge which is the percentage of battery that can be safely drained with any damage whereas, a lead-acid battery is cheaper.

3.3.2. Bidirectional converter

A bidirectional converter (BDC) shown by figure 3.8 is responsible for suppling power to battery that is battery charging and, delivering power to loads from the battery in case of high-power demands that is battery discharging. A dc to dc bidirectional converter is employed as a crucial component for connecting battery between source and load in a renewable energy system for un-interrupted power flow since the output of a renewable energy system swings due to conditional changes in weather. Bidirectional dc to dc converters may govern the flow of power in both directions between two dc sources

and loads by employing a precise switching technique. During high power demand, limited PV power availability, or both, a DC-DC converter is used to transport power from the battery to the load. When power consumption is low and available PV power is high, the obtained solar power can be supplied to the battery which can be utilised in high power demand conditions.

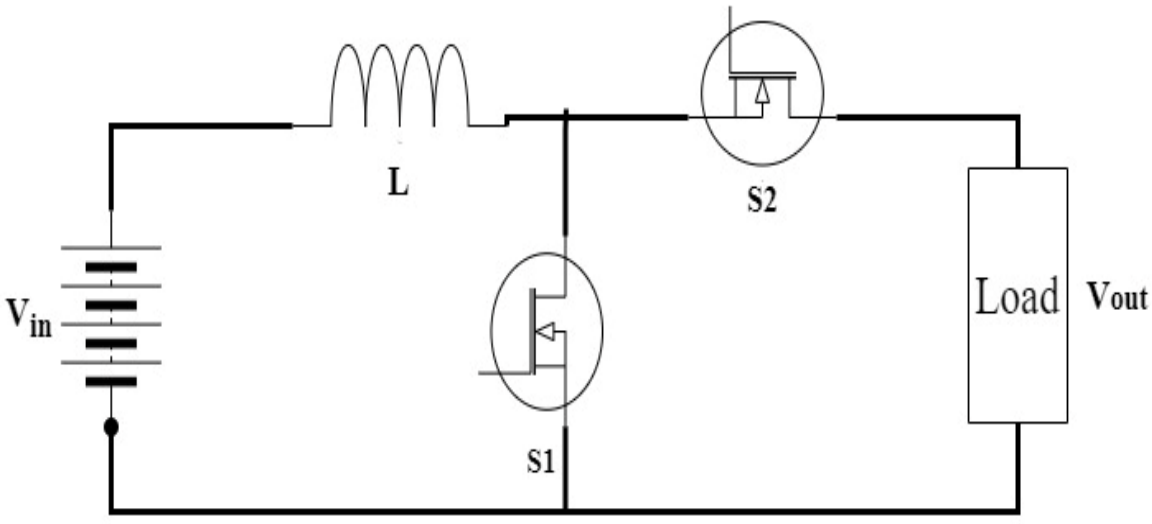


Fig 3.8: Model of Bidirectional converter

The BDC is a bidirectional buck-boost converter that when sending power from the battery to the load acts as a boost converter, and when sending power from DC bus to battery it acts as a buck converter. Figure 3.9 depicts the BDC in its two modes of operation, according to figure 3.9(a) the BDC works in boost mode where the battery acts as a source and the DC bus acts as a sink. The power is delivered by the battery to help maintain the DC bus voltage this is achieved when the switch S1 behaves as a switch and S2 behaves as a diode. This condition can be observed in conditions when the load demand is more that the obtainable PV power. According to figure 3.9(b) the BDC behaves as a buck converter where, the DC bus acts as the source and the battery behaves as a sink. This facilitates in charging the battery by operating S2 as a switch and S1 as diode. This condition can be observed when the load demand is low and available PV power is high.

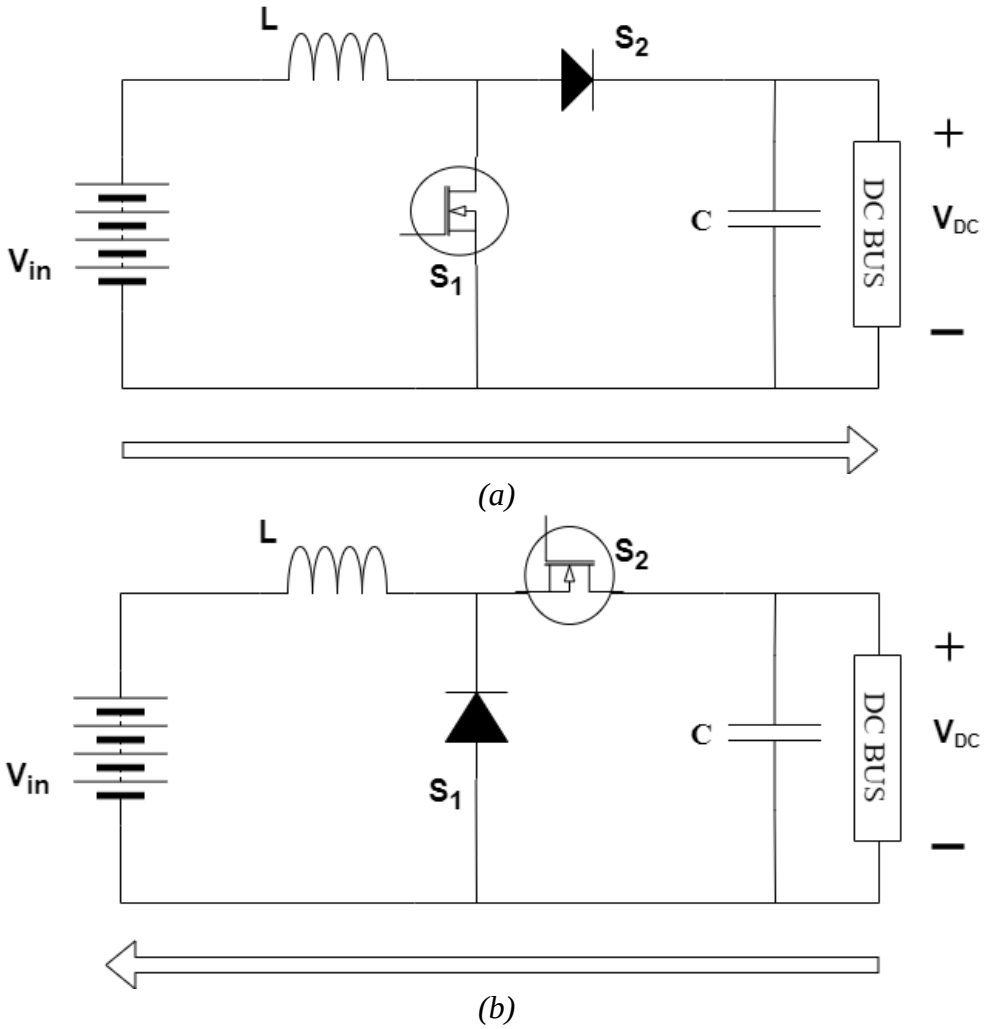


Fig 3.9: Operation of bidirectional converter (a) BDC as boost (b) BDC as buck

For a smooth operation between charging and discharging of battery suitable inductor and capacitor has to be selected according to the given formula,

While operating the BDC in buck mode,

$$L_{buck} > \frac{V_{out}}{2I_{out}f} \quad , C_{buck} = \frac{2I_{out}}{8f\Delta V_o} \tag{3.16}$$

Here V_{out}, I_{out} can be replaced by V_{in}, I_{in} which is the battery voltage and current respectively.

While operating the BDC in boost mode,

$$L_{boost} > \frac{d(1-d)^2R}{2f} \quad , C_{boost} = \frac{dV_{out}}{Rf\Delta V_o} \tag{3.17}$$

Here R represents the maximum load that can be applied at the DC bus, d is the duty ratio and f represents the switching frequency.

$$L = \max(L_{buck}, L_{boost}) \ , \ C = \max\left(C_{buck}, C_{boost}\right) \tag{3.18}$$

3.4. LOADS

Loads that are applied to the DC bus can be of two types, DC loads or AC loads. DC loads can be coupled straight with the DC bus whereas, AC loads involve an inverter to converter the DC power from the DC bus into AC power. As shown by figure 3.10, DC loads are directly coupled to the DC bus but, AC loads require an inverter between DC bus and AC loads.

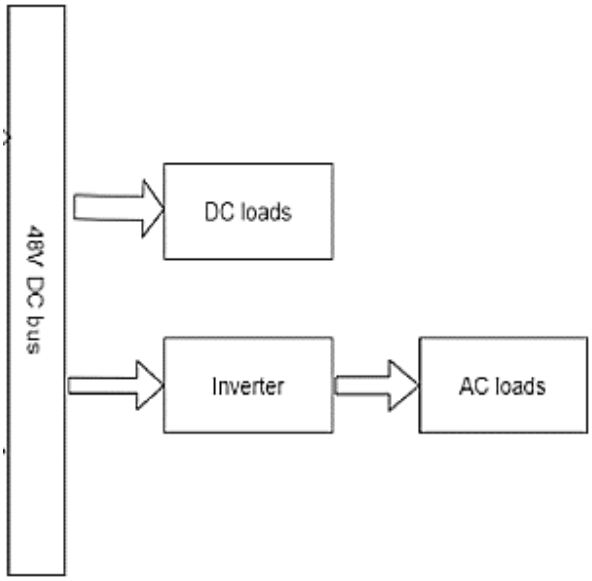


Fig 3.10: Coupling of DC and AC loads to the DC bus.

3.4.1. AC Loads

AC loads are coupled to DC bus through an inverter which is a DC-AC converter that manages conversion of DC power into an AC power by using various switching devices. The AC loads used with the SPMS are rated for household applications thus, the required AC voltage is 220V at 50Hz since, the DC bus used in the system is rated at 48V so there is a need for voltage boost at in intermediate stage to match the output AC voltage rating. The AC subsystem is design in such a way that it acts in two stages of operation, stage 1 represents the DC boost stage and stage 2 represents the inverter stage as shown in figure 3.11.

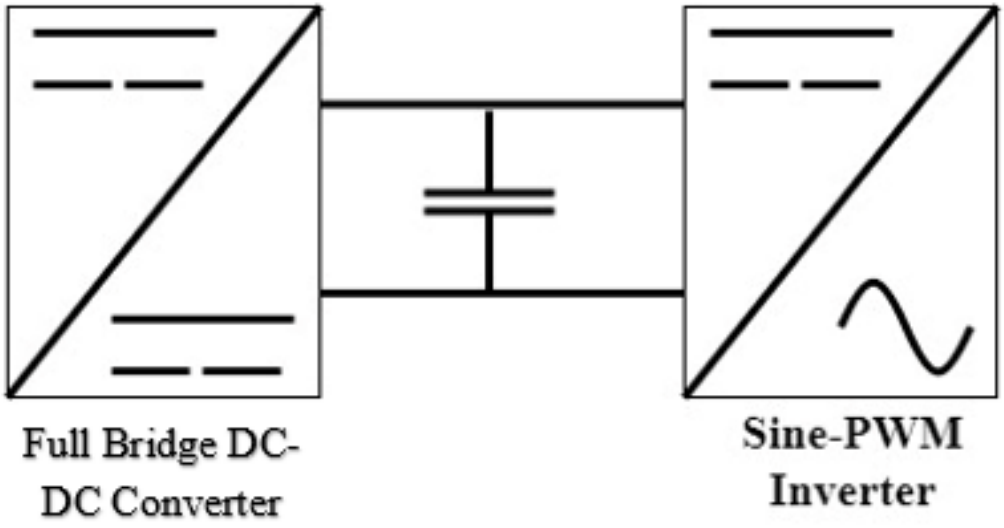


Fig 3.11: Two stages of Inverter system

3.4.1.1. DC boost Stage

In the stage 1, the DC bus voltage is boosted to a greater voltage by the means of a DC-DC converter. In this a full bridge DC-DC converter is employed as it provides an isolation between the high voltage side and low voltage side by the means of a transformer. Figure 3.12, represents a full bridge converter that is designed as a voltage boost converter, this is achieved by the presence of a transformer that helps in transforming the voltage magnitude and also providing an isolation between input side and output side.

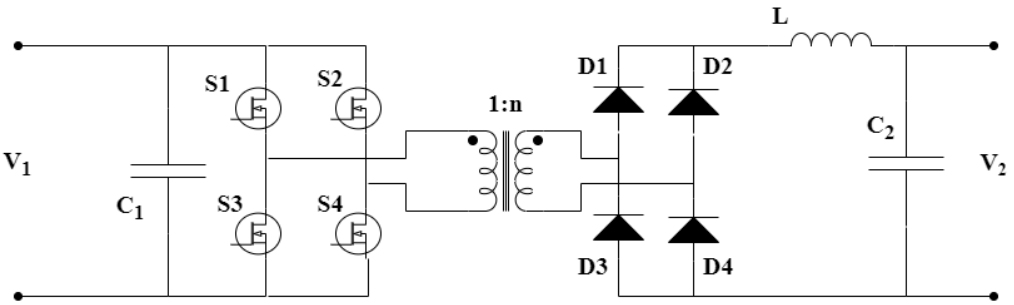


Fig 3.12: Full bridge DC-DC converter

On the input side of converter, a full bridge configuration is used which consists of switching devices S1, S2, S3 & S4. These switching devices are switched a high frequency to allow an AC voltage build up on the primary of the transformer. Switch S1 & S4 are switch ON together allowing the current to flow in one direction whereas, when switches S2 & S3 are switch ON the current flows in other direction, thus producing an AC current and voltage. This primary side voltage is reflected on the secondary of transformer as a high frequency voltage which is rectified by the full bridge diode

configuration at the secondary. Depending upon the transformation ratio of the transformer the voltage at the secondary is step up version of input voltage. To reduce the high frequency ripples at the secondary output terminals a low pass filter is applied using inductor (L) and capacitor (C_2).

Applying small signal analysis on figure 3.13 to linearize the non-linear system in order to analyse the switching converter. To analyse the converter some assumptions, have to be made such as the converter is operating in stable mode of operation and when a small disturbance is applied it can be roughly be viewed as a linear system.

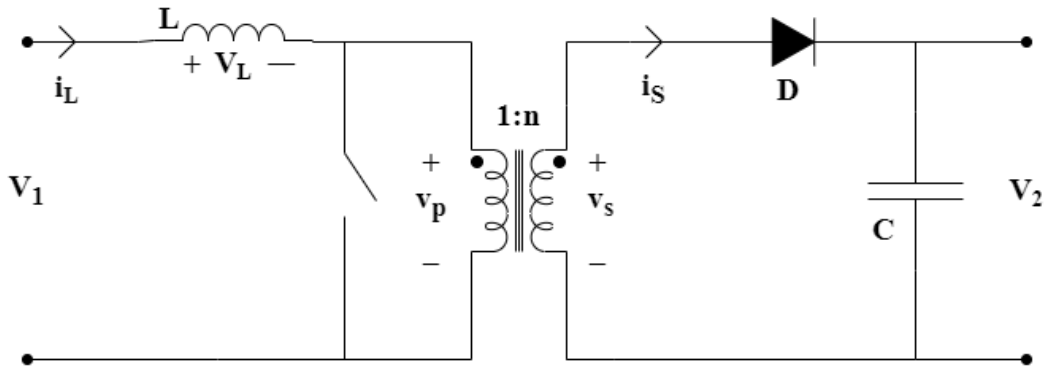


Fig 3.13: Equivalent circuit of full bridge converter in boost mode

For dynamic modelling of full bridge converter in boost mode, from electrical equations of capacitor current and inductor voltage i_c, V_L can be written as

$$V_L = \frac{Ldi(t)}{dt} \ , \ Ic = \frac{Cdv(t)}{dt} \tag{3.19}$$

Based on figure 3.13 following equations can be derived as part of small signal analysis,

$$\begin{cases} L\frac{d(<i_L(t)>_{T_s})}{dt} = -<v_1(t)> + \frac{<v_2(t)>}{n}(1-<d(t)>) \\ C\frac{d(<v_2(t)>_{T_s})}{dt} = (1-<d(t)>)<i_s(t)> - \frac{<v_{2(t)}>}{R} \\ <i_s(t)> = \frac{1}{n}<i_L(t)>(1-<d(t)>) \end{cases} \tag{3.20}$$

Assuming the system is working at steady state and small disturbances are added to d(t) and $v_1(t)$ resulting in $<d(t)> = D + \hat{d}(t), <v_1(t)> = v_1 + \widehat{V_1}(t)$

Now the equation 20 can be rewritten as,

$$\begin{cases} L\frac{d\widehat{i_L}(t)}{dt} = -\widehat{v_1}(t) + \frac{1-D}{n}\widehat{v_2}(t) + \frac{v_2}{n}\hat{d}(t) \\ C\frac{d\widehat{v_2}(t)}{dt} = (1-D)\widehat{i_s}(t) - I_s\hat{d}(t) - \frac{\widehat{v_2}(t)}{R} \\ \widehat{i_s}(t) = \frac{1-D}{n}\widehat{i_s}(t) - \frac{I_L}{n}\hat{d}(t) \end{cases} \tag{3.21}$$

Converting the equation 3.21 into s-domain,

$$\begin{cases} SL\widehat{i_L}(t) = -\widehat{v_1}(s) + \frac{1-D}{n}\widehat{v_2}(s) + \frac{v_2}{n}\hat{d}(s) \\ SC\widehat{v_2}(t) = (1-D)\widehat{i_s}(s) - I_s\hat{d}(s) - \frac{\widehat{v_2}(s)}{R} \\ \widehat{i_s}(s) = \frac{1-D}{n}\widehat{i_s}(s) - \frac{I_L}{n}\hat{d}(s) \end{cases} \quad (3.22)$$

The control to output transfer function of full bridge converter can be written as,

$$G_{vd}(s)|_{v_1(s)=0} = \frac{\widehat{v_2}(s)}{\hat{d}(s)} = V_2 \frac{1-\frac{n^2L}{R(1-D)^2}S}{LCnS^2+\frac{Ln}{R}S+\frac{(1-D)^2}{n}} \quad (3.23)$$

For design of controller to operate in closed loop mode we need to design a specific compensator in order to improve the stability of the entire system. For controller design, the open loop transfer function of the system can be specified as,

$$G(s)H(s) = G_c(s)G_m(s)G_{vd}(s)H(s) \quad (3.24)$$

The compensator designed for this system is a PI compensator thus the transfer function can be written as;

$$G_c(s) = k_P + \frac{K_I}{s} \quad (3.25)$$

Where K_P, K_I represent the proportional and integral constants. For the system to be stable the gain margin and phase margins must be positive or phase margin should be higher than gain margin.

3.4.1.2. Inverter Stage

The next stage in operating the AC loads is converting the boosted DC voltage into AC voltage by using an inverter. An inverter is a DC-AC converter that is capable to transforming the DC voltage into AC voltage by the means of switching devices. For this operation, a single-phase full bridge sinusoidal PWM (S-PWM) inverter was used where the modulation signal is a sinusoidal waveform and carrier signal is the triangular waveform. Figure 3.14 depicts the circuit diagram for a single-phase inverter with S-

PWM signal, this S-PWM signal is generated at a higher switching frequency which is determined based on the carrier frequency that is typically in kHz range.

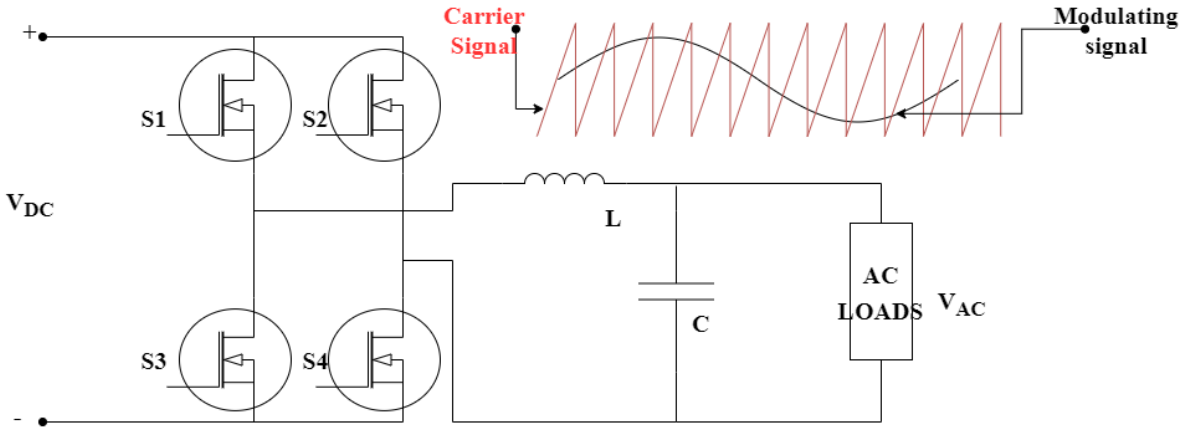


Fig 3.14: Full bridge single-phase inverter with sinusoidal PWM signal

Due to this an output at full bridge terminals is high frequency pulses in the range of kHz, so to reduce the output frequency additional LC filter is applied that removes the high frequency pulses to produce a 50Hz sinusoidal AC output.

$$L = \frac{R_L\sqrt{2}}{2\pi f_o}, \quad C = \frac{1}{2\pi f_o R_L\sqrt{2}} \tag{3.26}$$

Here R_L represents the load, f_o represents the cut-off frequency

3.4.2. DC loads

Loads that require DC power are termed as DC loads. These loads can be coupled to the DC bus with or, without any addition DC-DC converter in between. DC power can be utilized in operations such as for lighting loads, motoring loads or battery charging. For efficient utilization of electrical energy in lighting sector DC lighting offers higher efficiency and utilization, by dimming and occupancy sensing operations. Other examples of DC loads are brushless DC (BLDC) motors which are used as fans that provide high efficiency and low noise when operated. Such loads can be directly coupled to the DC bus or through a DC-DC converter to match the voltage ratings.

CHAPTER 4

DEEP REINFORCEMENT LEARNING BASED CONTROL

4.1. DEEP REINFORCEMENT LEARNING

Reinforcement learning is the process of teaching machine learning models to make a sequence of decisions. The agent learns to achieve a goal in an unpredictable, perhaps difficult environment. In reinforcement learning, an artificial intelligence encounters a game-like situation. The machine use trial and error to arrive at a solution. In order to induce artificial intelligence to do what the programmer wants; it is given either incentives or penalties for the acts it does. Its goal is to make the total prize as high as feasible. Despite the fact that the designer defines the reward policy–that is, the game's rules–he offers no advice or suggestions to the model on how to complete the game. It's up to the model to figure out how to accomplish the task in order to maximise the reward, starting with entirely random trials and escalating to complicated plans and superhuman skills. Reinforcement learning is presently the most effective technique to hint machine creativity by utilising the power of search and numerous trials.

4.1.1. Q-learning

Considering the agent interacts with the environment E, at time t and makes an observation x_t where it takes an action at time t thus, receiving a reward r_t. The entire environment can be mapped into state, action pairs as $s_t = (x_1, a_1 \ldots\ldots a_{t+1}, x_t)$. The actions performed by the agent are based upon the policy π, which correlates the states probability distribution with actions as $\pi: S \rightarrow p(A)$ thus modelling the state space S, action space A, with initial probability distribution $p(S_1)$, transaction dynamics

$p(s_{t+1}|s_t, a_t)$ and reward $r(s_t, a_t)$ into a Markov decision process. The output for states can be defined as the summation of upcoming discounted rewards,

$R_t = \sum_{i=t}^{T} \gamma^{i-t} r(s_i, a_i)$ where ɣ represents discount factor. The action-value function relates the expected return after an action is taken at state s_t. and can be written as,

$$Q^{\pi}(s_t, a_t) = E_{r_t, s_{t+1} \sim E}[r(s_t, a_t) + \gamma E_{a_{t+1} \sim \pi}[Q^{\pi}(s_{t+1}, a_{t+1})]] \quad (4.1)$$

If the target policy is deterministic then we can rewrite the inner expectation as,

$$Q^{\mu}(s_t, a_t) = E_{r_t, s_{t+1} \sim E}[r(s_t, a_t) + \gamma[Q^{\mu}(s_{t+1}, \mu(s_{t+1}))]] \quad (4.2)$$

A Q-table is used to maintain the learning progress in Q-learning methods. As the number of state-action pairs grows, so does the size of the Q-table, and the process suffers from the curse of dimensionality. As a result, traditional Q-learning algorithms require significant discretization stages, resulting in unsatisfactory outcomes. Furthermore, because the Q-function is deterministically computed, Q-learning cannot handle stochastic policies.

4.1.2. Deep Deterministic Policy Gradients (DDPG)

DDPG is a deep reinforcement learning based algorithm in which a Q-network present in RL is replaced by a neural network thus making it capable of having multiple state, action pairs while simultaneously reducing the look up time as associated with RL based look-up tables, the architecture of DDPG algorithm can be observed from figure 4.1. A Q network, a deterministic policy network, a target Q network, and a target policy network are used in DDPG. The Q network and policy network are similar to basic Advantage Actor-Critic, except that in DDPG, the Actor directly correlates the states to actions (the network's output is directly the output) rather than producing the probability distribution across a discrete action space. The learnt networks are slowly tracked by the target networks, which are delayed time duplicates of their initial networks. The use of these target value networks considerably improves learning stability. Policy gradient can be expressed as:

$$\nabla_{\theta} J(\theta) = E_T[\sum_{t=0}^{T-1} \nabla_{\theta} log \pi_{\theta}(a_t|s_t) G_t] \quad (4.3)$$

Decomposing the expectation results in

$$\nabla_{\theta} J(\theta) = E_T[\sum_{t=0}^{T-1} \nabla_{\theta} log \pi_{\theta}(a_t|s_t) Q_w(s_t, a_t)] \quad (4.4)$$

Where s_t, a_t are the state and action of the network.

A replay buffer is also used by DDPG to collect experience and update neural network. It consists the experience of any agent which is a tuple consisting of (state, action, reward,

next state) that is stored as a finite-sized throughout each trajectory roll-out. After the roll out from the replay buffer takes place, the updated Q-value is obtained as

$$y_i = r_i + \gamma Q'\left(S_{i+1}, \mu'\left(S_{i+1}\middle|\theta^{\mu'}\right)\middle|\theta^{Q'}\right) \quad (4.5)$$

where γ is the discount factor. The target value network and target policy network are used to determine the next-state Q values. The mean-squared loss between the modified Q value and the original Q value is then minimised. The objective here is to maximise the expected return of the policy function $\nabla_\theta J(\theta)$. For continuous action spaces, exploration is done by adding noise N to the action this can be expressed as

$$\mu'(s_t) = \mu\left(s_t\middle|\theta_t^{\mu}\right) + N \quad (4.6)$$

4.2. DDPG ALGORITHM

Set the critic $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ randomly.

Set target network Q' and μ' with weights $\theta^Q \rightarrow \theta^{Q'}$ and $\theta^\mu \rightarrow \theta^{\mu'}$

Set replay buffer *R*

for episode = 1, M **do**

Set a random process N for exploring actions

Obtain initial observation state s_1

for t = 1, T **do**

Pick action $at = \mu(s_t|\theta^\mu) + N$ according to current policy and exploration noise

Run action a_t , receive reward r_t and observe state s_{t+1}

Capture transition (s_t, a_t, r_t, s_{t+1}) in R

Sample random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R

Set $y_i = r_i + \gamma Q'\left(S_{i+1}, \mu'\left(S_{i+1}\middle|\theta^{\mu'}\right)\middle|\theta^{Q'}\right)$

Update critic by minimizing the loss:

$$L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$$

Update the actor policy:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} (s|\theta^\mu)|_{s_i}$$

Update the target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

end for

end for

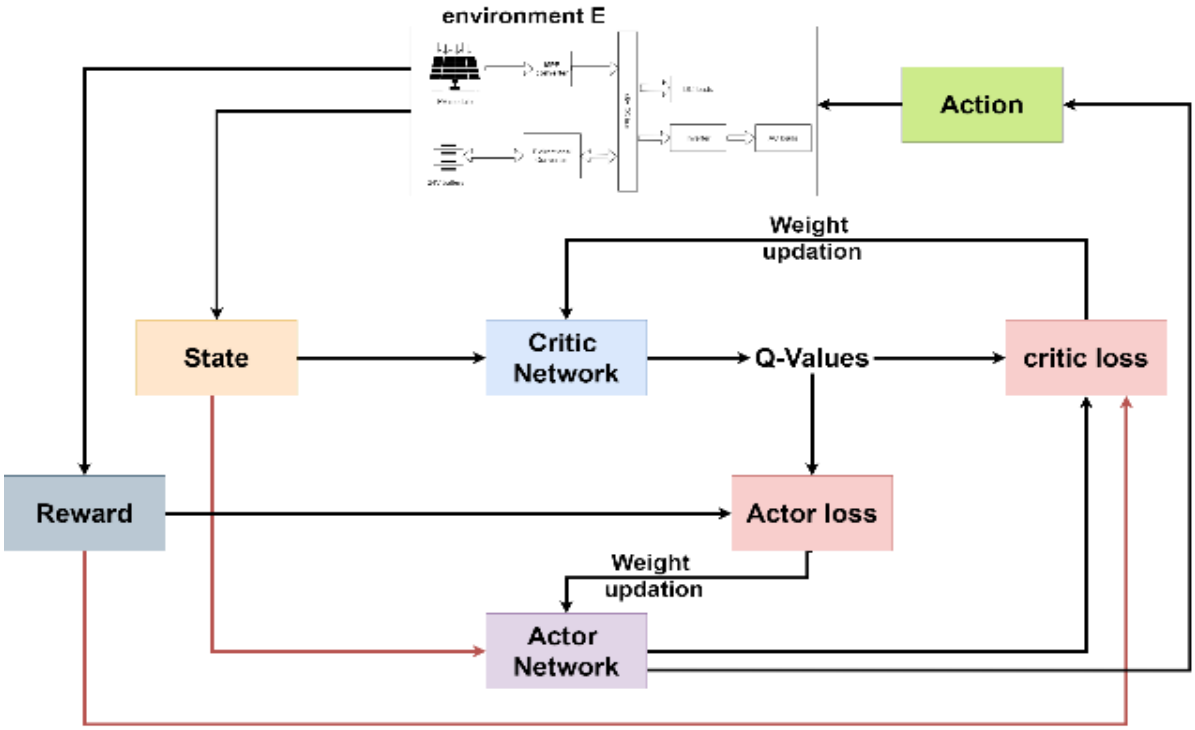


Fig 4.1: Architecture of DDPG agent

CHAPTER 5

SIMULATION SETUP & CALCULATIONS

The simulation for the scope of this project is simulated in MATLAB (Simulink), using which the entire PV based system was designed and analysed. For the purpose to developing and training of DRL agent Reinforcement Learning Toolbox and MATLAB (Simulink) were used respectively.

5.1. DESIGNING OF SYSTEMS

The various components of the PV system are designed in MATLAB (Simulink) based upon the design specifications of the system. According to the system requirements a PV battery-based system is to be designed with 800W solar PV array connected to a 48V DC bus, the system also includes an energy storage system with a 24V battery system capable of delivering constant power of 800W to the DC bus for a duration of 7 hours continuously without operating the PV system.

Along with DC loads the PV battery-based system has to operate AC loads as well, so an additional inverter has to be designed that is rated at 220V, 50Hz. According to the design specifications, each subsystem of the PV battery-based system has to be designed. The components of PV-battery based system are solar power tracking system which consists of PV array and MPP converter, energy storage system comprising of a battery and bidirectional converter and loads that can be coupled to the DC bus.

5.1.1 Design of Solar Power Tracking System

As per the design specifications, the solar power tracking system has to be designed for 800W PV array and accordingly a maximum power point tracking converter has to be designed.

5.1.1.1 PV Array Design

For the design of PV array, the single module rating is specified at 213.15W at 1000W/m^2 irradiance with an open circuit voltage of 25V. According to the design specifications, to obtain 800W several PV modules have to be coupled together in series or parallel to get the desired power rating. By connecting two PV modules in series the obtained power is 426.3W, connecting 2 such strings in parallel results in 852.6W of solar power at irradiance of 1000W/m^2. Since 2 PV modules are connected in series thus, the open circuit voltage of the PV array equals 50V. Figure 5.1 shows the power outputs available at 100,1000 W/m^2, assuming that the minimum possible irradiance in a day is 100W/m^2.

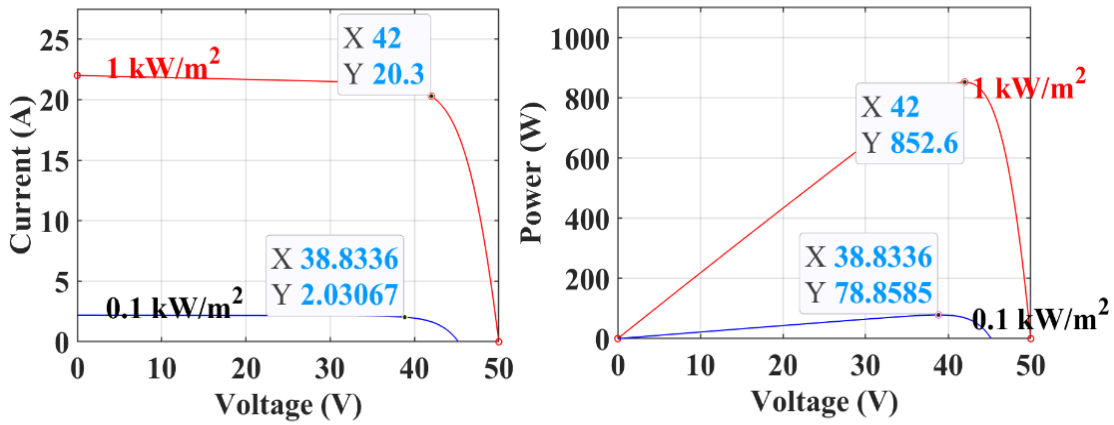


Fig 5.1: I-V & P-V characteristics of PV array

5.1.1.2 MPP Converter Design

For designing the MPP converter several parameters have to be considered such as input voltage (V_{in}), maximum load that can be applied (P_{max}), switching frequency (f) and duty ratio (d). The main elements to be designed for the MPP converter is the inductor (L_{boost}) and capacitor (C_{boost}) according to the equations 3.14 and 3.17.

For the design of MPP converter at a switching frequency (f) of 50khz and maximum power capability (P_{max}) of 852.6W at 1000W/m^2 and 78.85W at 100W/m^2, as the MPP converter is coupled to a 48V DC bus the output voltage (V_o) can be considered as 48V.

As per the assumption that the minimum and maximum irradiance available in a day is 100W/m^2 and 1000W/m^2 respectively therefore, according to the figure 16 the minimum and maximum input voltage (V_{in}) are 38.83V & 42V respectively. Considering equation 3.13,

$$Minimum\ duty\ ratio\ (dmin)\ =\ 1-\frac{42}{48}=0.125 \quad (5.1)$$

$$Maximum\ duty\ ratio\ (dmax)\ =\ 1-\frac{38.83}{48}=0.191 \quad (5.2)$$

Assuming the minimum load (P_{min}) that can be attached to the system is 12W,

$$R_{load}=\frac{V_o^2}{P_{MIN}}=\frac{48^2}{12}=200\Omega \quad (5.3)$$

From equations 3.14, 5.1 & 5.3,

$$L_{boost}>\max\left(\frac{0.125(1-0.125)^2\times200}{2\times50\times10^3}\ ,\frac{0.191(1-0.191)^2\times200}{2\times50\times10^3}\right) \quad (5.4)$$

$$L_{boost}>\max(191.4\mu H, 250\mu H) \quad (5.5)$$

Similarly, C_{boost} can be calculated as, according to equation 3.17,

$$C_{boost}=\frac{0.191\times48}{2\times50\times10^3\times0.05}=1833.6\mu F \quad (5.6)$$

For practical purpose considering the L_{boost}, C_{boost} as 375µH & 3000µF.

5.1.2 Design of Energy Storage System

The energy storage system consists of 2 components mainly a battery and bidirectional converter, according to the design specifications the battery should be proficient in keeping the load power of 800W for a maximum of 7 hours while keeping the DC bus voltage constant.

5.1.2.1 Battery selection

As Lithium-ion batteries possess superior charging and discharging capabilities to that of lead acid batteries, for this system design Li-ion battery is considered. A single Li-ion cell is rated at 3.7V nominal voltage with 2500mAh capacity. According to the design specifications, the battery should be able to handle 800W load power for 7 hours continuously so, considering equation 3.15,

$$Required\ capacity\ =\ \frac{800\times7}{24}=\ 233.33\ Ah \quad (5.7)$$

For design purpose battery capacity can be selected as 250Ah. The battery system has to be designed for 24V thus, individual cells have to be connected in series or parallel to

meet the required voltage and capacity ratings. Nominal voltage (V_{norm}) of the 24V battery is 20.54V whereas, individual cell nominal voltage (V_{cell_norm}) is 3.7V.
To achieve the battery specifications,

$$Series\ connected\ cells = \frac{V_{norm}}{V_{cell-norm}} = \frac{20.54}{3.7} = 5.55\ \approx 6 \quad (5.8)$$

$$Series\ strings\ in\ parallel = \frac{Battery\ capacity}{cell\ capacity} = \frac{250Ah}{2500mAh} = 100 \quad (5.9)$$

We require a total of 600 cells to form the battery rated for 24V, 250Ah.

5.1.2.2 Bidirectional Converter Design

The bidirectional converter is responsible for charging and discharging the battery while maintaining a constant DC bus voltage, this is achieved by operating the BDC in buck or boost mode respectively. While designing the BDC different operating modes has to be considered.

Design of BDC in buck mode, in this mode the BDC is responsible for battery charging so, the battery voltage can be considered as the output voltage (V_o) and the DC bus voltage can be considered as V_{in}. Assuming the minimum (P_{min}) and maximum (P_{max}) power available for battery charging is 50W and 852.6W respectively while the switching frequency is 50kHz. Considering equation 3.15,

$$L_{buck} = \frac{V_o}{2\left(\frac{Pmin}{V_o}\right)f} = \frac{24}{2\left(\frac{50}{24}\right)(50\times10^3)} = 115.2\mu H\ , \quad (5.10)$$

$$C_{buck} = \frac{2\left(\frac{P_{MAX}}{V_o}\right)}{8f\Delta v_o} = \frac{2\times\frac{852.6}{24}}{8\times50\times10^3\times0.05} = 3552.5\mu F \quad (5.11)$$

Design of BDC in boost mode, In this mode the BDC discharges the battery to maintain a constant DC bus voltage therefore, we can consider the DC bus voltage as V_o at 48V whereas the 24V battery voltage is treated as V_{in}. Assuming the minimum (P_{min}) and maximum (P_{max}) power required to maintain a constant DC bus voltage is 10W and 1000W respectively.
Considering equation 3.17,

$$L_{boost} = \frac{d(1-d)^2V_o^2}{2fP_{MIN}} = \frac{0.5(1-0.5)^2 48^2}{2\times50\times10^3\times10} = 288\mu H \quad (5.12)$$

$$C_{boost} = \frac{dV_o}{\left(\frac{Vo^2}{P_{MAX}}\right)f\Delta V_o} = \frac{0.5\times48}{\left(\frac{48^2}{1000}\right)50\times10^3\times0.05} = 4166.67\mu F \quad (5.13)$$

From equations 5.10 & 5.12,

$$L = max(L_{buck}, L_{boost}) = 288\mu H \tag{5.14}$$

From equations 5.11 & 5.13,

$$C = \max(C_{buck}, C_{boost}) = 4166.67\mu F \tag{5.15}$$

For design purpose L = 450μH, C = 6000μF.

5.1.3 Connection to AC Loads

Connection of AC loads to a DC bus requires an inverter in between them which converts the DC from the DC bus into AC. For the scope of this project a two-stage approach was employed in which first the DC bus voltage was stepped-up to a higher 390V then applied to a SPWM based inverter to converter the 390V DC into 220V(rms), 50Hz AC voltage.

5.1.3.1 Design of Full Bridge Converter

The full bridge converter is considered as the first stage in operating AC loads where the DC bus voltage is boosted to a higher DC voltage level for operating the inverter. Since the DC bus voltage has to be boosted to 390V therefore, V_{in} can be considered to be 48V and V_{out} as 390V

Considering a transformer ratio ($n = \frac{N_2}{N_1}$) of 10,

$$d = \frac{V_o}{2V_{in}n} = \frac{390}{2\times48\times10} = 0.40625 \tag{5.16}$$

For controller design the transfer function for full bridge DC-DC converter has to be analysed with inductor as $800\mu H$ and capacitor as $3000\mu F$, considering equation 3.23:

$$G_{vd}(s)|_{v_1(s)=0} = \frac{\left(1-\frac{n^2L}{R(1-D)^2}S\right)V_o}{LCnS^2+\frac{Ln}{R}S+\frac{(1-D)^2}{n}} = \frac{390-0.867s}{2.4\times10^{-4}s^2+8\times10^{-6}s+0.036} \tag{5.17}$$

According to figure 5.2, the phase margin of $G_{vd}(S)$ is negative thus, making the system unstable.

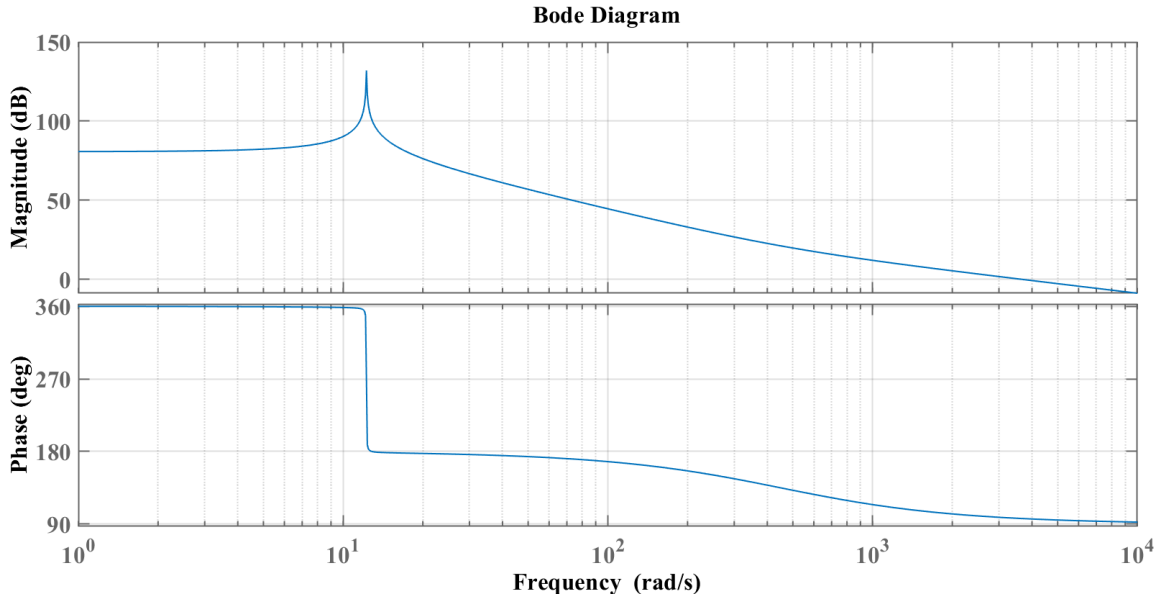


Fig 5.2: Bode plot for $G_{vd}(s)$

To improve upon the stability a compensator has to be analysed, the compensation used in this system is PI controller whose transfer function can be given by the equation 3.25:

$$G_c(s) = \frac{K_p s + K_i}{s} \tag{5.18}$$

Where K_p, K_I are 1×10^{-7} *and* 1×10^{-6} respectively.

From equation 3.24, 5.17 & 5.18,

$$G(s)H(s) = \frac{-8.67\times10^{-8}s^2 + 3.813\times10^{-5} + 0.00039}{0.00024s^3 + 3.813\times10^{-6}s^2 + 0.036s} \tag{5.19}$$

G(s)H(s) represents the open loop transfer function where H(s) is 1. Figure 5.3 represents the bode plot for open loop transfer function with stable operation as gain margin and phase margin for the system are positive.

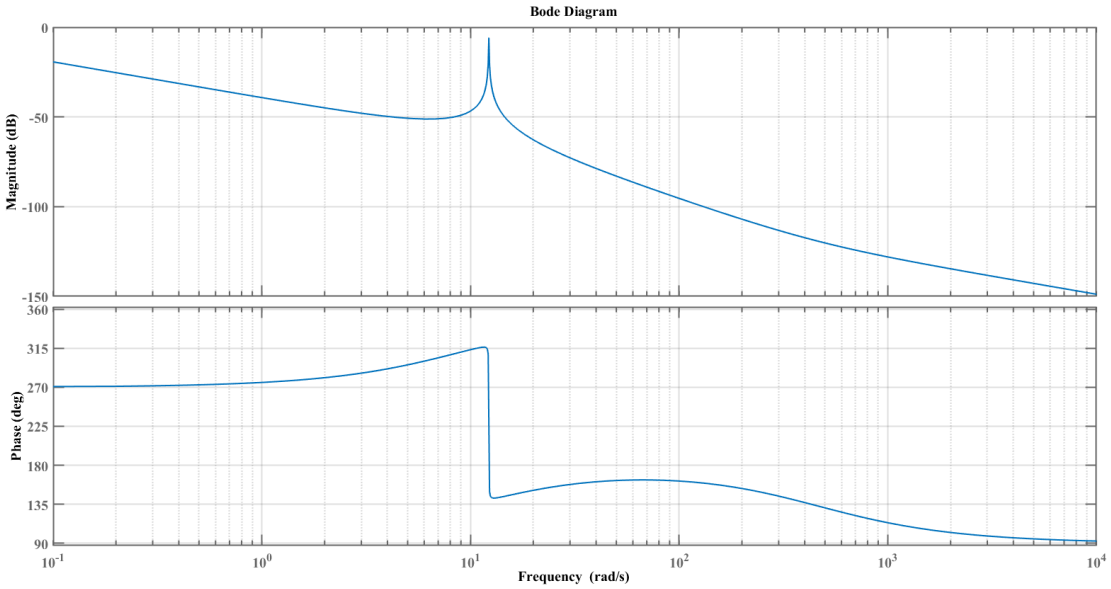


Fig 5.3: Bode plot for G(s)H(s)

5.1.3.2 Design on SPWM inverter

The inverter is employed for converting the DC voltage to an AC voltage of fixed magnitude, for this a sinusoidal PWM inverter technique was used. The modulating signal is a sinusoidal wave and the carrier is a triangular wave with modulation index (m_i) of 0.88. The fundamental frequency component of output voltage (V_{ab}) can be given by,

$$V_{ab} = m_i V_{DC} = 0.88 \times 390 = 344V \quad (5.20)$$

5.2 DDPG ARCHITECTURE

For development for actor-critic network, a fully connected dense layer was selected for interaction with environment and perform the tasks. The actor network consists of five fully connected dense layers with 1000 neurons each and tanh activation function. These layers are bifurcated to add the complexities that are observed by the MPP and BDC control as shown in figure 5.4(a). whereas, the critic network consists of state path and action path, for state path three fully connected layers were used with 500 neurons each and consisted of ReLU activation function and action path consisted of three FC layer with 500 neurons each as shown by figure 5.4(b). Two deep neural networks are used in the actor-critic (AC) technique. The policy network (or the actor) receives input from the environment and performs a control action (or a policy). The estimator for the value function (or the critic) evaluates the condition of the environment and the reward

acquired from the actor's control action before returning the action's estimated value. According to the policy gradient theorem, the actor-network uses a gradient ascent technique to maximise the objective function ∇_θ (eqn 4.4). To estimate the value function, the critic network utilises gradient descent to reduce error.

(a) (b)

Fig 5.4: structure of DDPG agent (a) actor network (b) critic network

5.3 SIMULATION SETUP

The following sections explain the design aspects of SPMS in MATLAB(Simulink), the sub-systems of Simulink model are categorised as simulation of PV system and simulation of DRL-agent. The simulation of PV system describes the electrical components of SPMS whereas, the simulation of DRL-agent describes the design and modelling of deep reinforcement learning agent for calculating reward and observations.

5.3.1 Simulation of PV system

As analysed in previous sections the PV sub-system has been designed with the PV array connected to a boost converter for MPP tracking, at various irradiance conditions. The PV sub-system also consists of a 24V, 250Ah battery and BDC converter for battery charge and discharge operations when load demand is low and high respectively depicted by figure 5.5. The BDC acts a buck converter during charging of the battery carrying power from PV array to battery and works as a boost converter in situations of high load demand hence providing power to the load alongside the PV array.

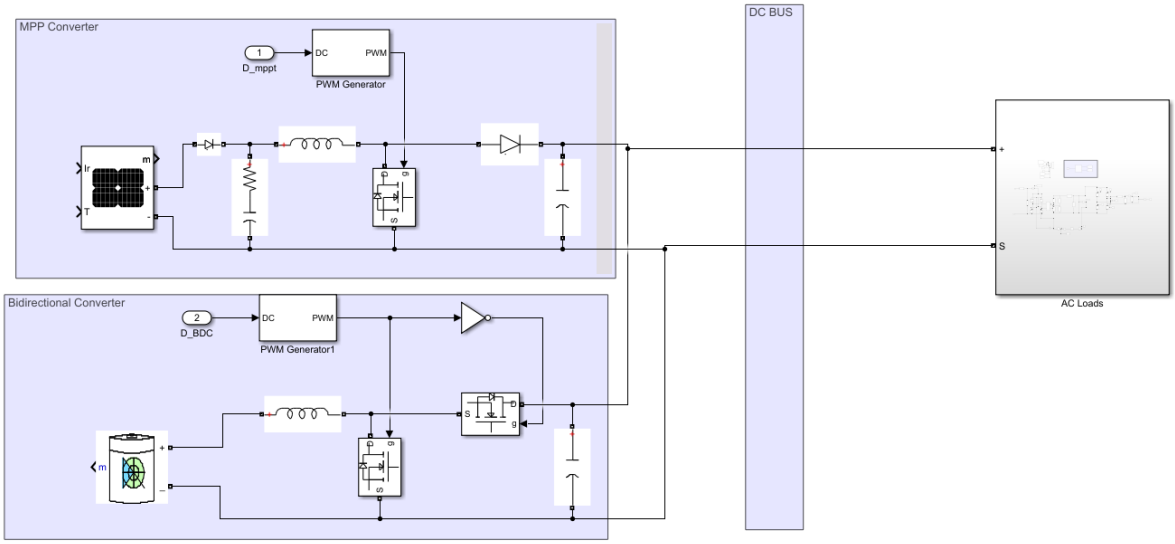


Fig 5.5: MATLAB-simulink model for PV sub-system

For design parameter specifications the values for each parameter used for various sub-systems is defined in table 5.1.

Unit	Parameter	Value
MPP converter	Inductor (L)	375μH
	Capacitor (C)	3000μF
	Switching frequency (f)	50khz
BDC	Inductor (L)	450μH
	Capacitor (C)	6000μF
	Switching frequency (f)	50khz
PV Array	856.2W@1000W/m^2	
Battery	24V,250Ah	
Full Bridge Converter	Inductor (L)	800μH
	Capacitor (C)	3000μF
	Switching frequency (f)	50khz

Table 5.1: List of parameters used in PV system Design

5.3.2 Simulating Deep Reinforcement Learning Agent

While setting up simulations for DRL agents two most important aspects are rewards and observations, based on these the agent interacts and learns from the environment.

5.3.2.1 Training of DRL Agent

Before any DRL agent can be employed into use it is necessary to train the agent where, it learns about the available states and actions by mapping out the environment. For the purpose of developing an environment parameter such as action space and state space are defined, the action space determines the available actions that a DRL agent can take while state space represents the available states that the agent can discover based on explorations. The environment is generated using the code given in figure 5.6 which defines the observation space and the action space. Table 5.2 defines the parameters defined while training of DRL agent.

```
 %Defining observation space
obsInfo = rlNumericSpec([2 1],'LowerLimit',[-10 -10]','UpperLimit',[10 10]');
obsInfo.Name = 'observations';
obsInfo.Description = 'Output DC bus Voltage error, V_pv';
numObservations = obsInfo.Dimension(1);

%defining action space
actInfo = rlNumericSpec([2 1],'LowerLimit',[0.1 0.47]','UpperLimit',[0.22 0.65]');
actInfo.Name = 'Action';
actInfo.Description = 'duty for mppt, Duty for BDC';
numActions = actInfo.Dimension(1);

%creating enviroment
env = rlSimulinkEnv('Deep_MPPT_','Deep_MPPT_/RL Agent',obsInfo,actInfo);

%env reset Function

env.ResetFcn = @(in)localResetFcn(in);
```

Fig 5.6: MATLAB code for generating the environment

Parameter	Value
Training Episodes	5000
Critic learning rate	$1x10^{-2}$
Actor learning rate	$1x10^{-3}$
Min batch size	64
Experience buffer	$1x10^{6}$
Discount factor	0.99
State space lower limit	[-10 -10]
State space upper limit	[10 10]
Action space lower limit	[0.1 0.47]
Action space upper limit	[0.22 0.65]

Table 5.2: List of parameters for training DRL agent

5.3.2.2 Reward Design

The instantaneous reward that is given to the agent for performing any action in the environment based upon this reward the agent decides which of the performed action for any given state is yields maximum reward. This is calculated as

$$r = -\alpha(V_{ter} - V_{load}) - \beta(P_{pv-prev} - P_{pv}) \quad (5.21)$$

where $P_{pv\text{-}prev}$ is the solar panel output power that the panel has produced in

$t-1$ timestep, P_{pv} represents the current solar power output at time t, α represents the penalty factor for DC bus voltage, β represents the penalty factor for moving away from maximum power point and V_{ter} is the desired bus voltage. According to figure 5.6, the reward is calculated according to equation 5.21 where α & β are 20 and 5 respectively.

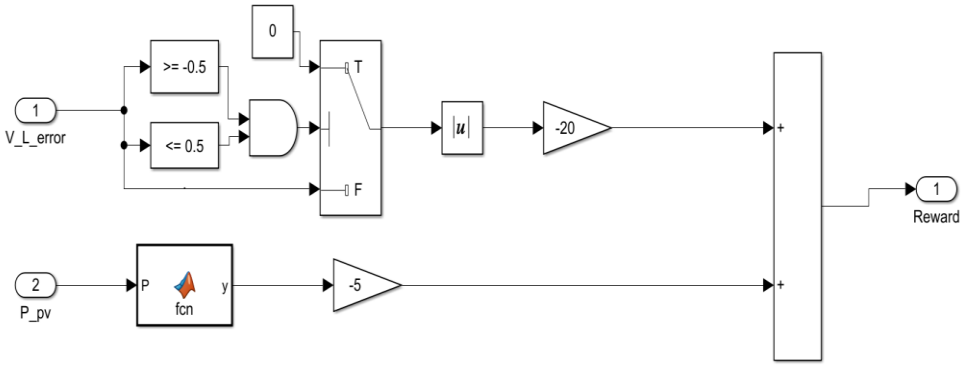


Fig 5.7: Reward for the DRL Agent

5.3.2.3 DRL Observation

Observation block shown by figure 5.7, is used for the interaction of RL agent with the PV subsystem where, the agent observes for changes in state after taking certain action.

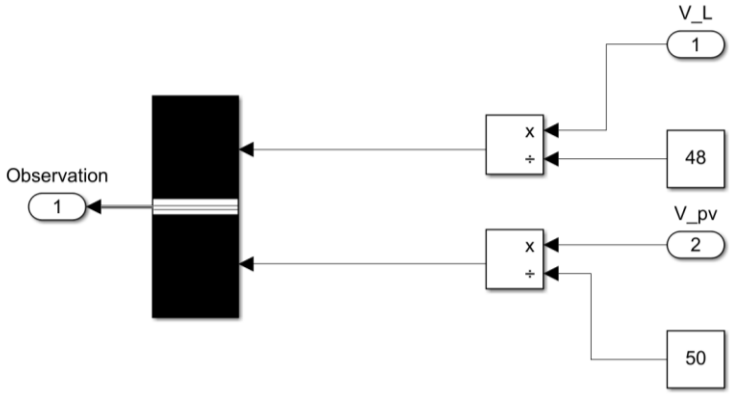


Fig 5.8: Observation block connected to DRL agent

The observation (state) is given as:

$$s = \{|V_{load}|', |V_{pv}|'\} \tag{5.22}$$

where $V'_{load} = \frac{V_{load}}{V_{ter}}$ & $V'_{pv} = \frac{V_{pv}}{V_{oc-pv}}$ and $V_{oc\text{-}pv}$ represents the open circuit PV voltage.

CHAPTER 6

RESULTS AND OBSERVATIONS

The following chapter describes the results to various simulation scenarios in which the solar power management system is operated. Depending upon the simulations, observations can be made upon the performance of deep reinforcement learning agent. The simulation scenarios are described in table 6.1 based on which the performance of DRL based SPMS is analysed.

Case	Irradiance (W/m^2)	Load	Load rating (W)
1	100 (constant)	DC (constant)	500
2	600 (constant)	DC (constant)	500
3	800 (constant)	DC (constant)	50
4	300 & 600 (Variable)	AC (constant)	1200
5	300, 600 & 800 (Variable)	DC (Variable)	50 to 500

Table 6.1: Conditions at which simulations are performed

6.1. CASE 1

In this case, the simulation is performed at 100W/m^2 irradiance. According to the P-V characteristics of the PV array as shown in figure 6.1, the maximum operating point power at 100W/m^2 is 78.85W.

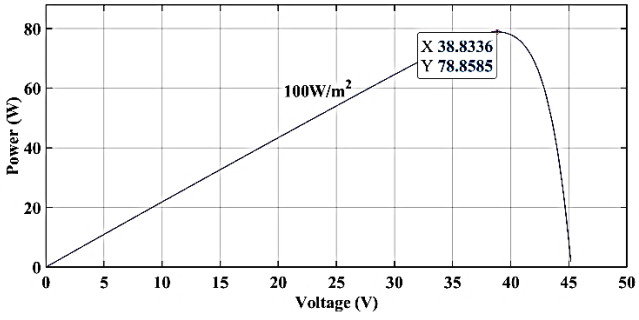
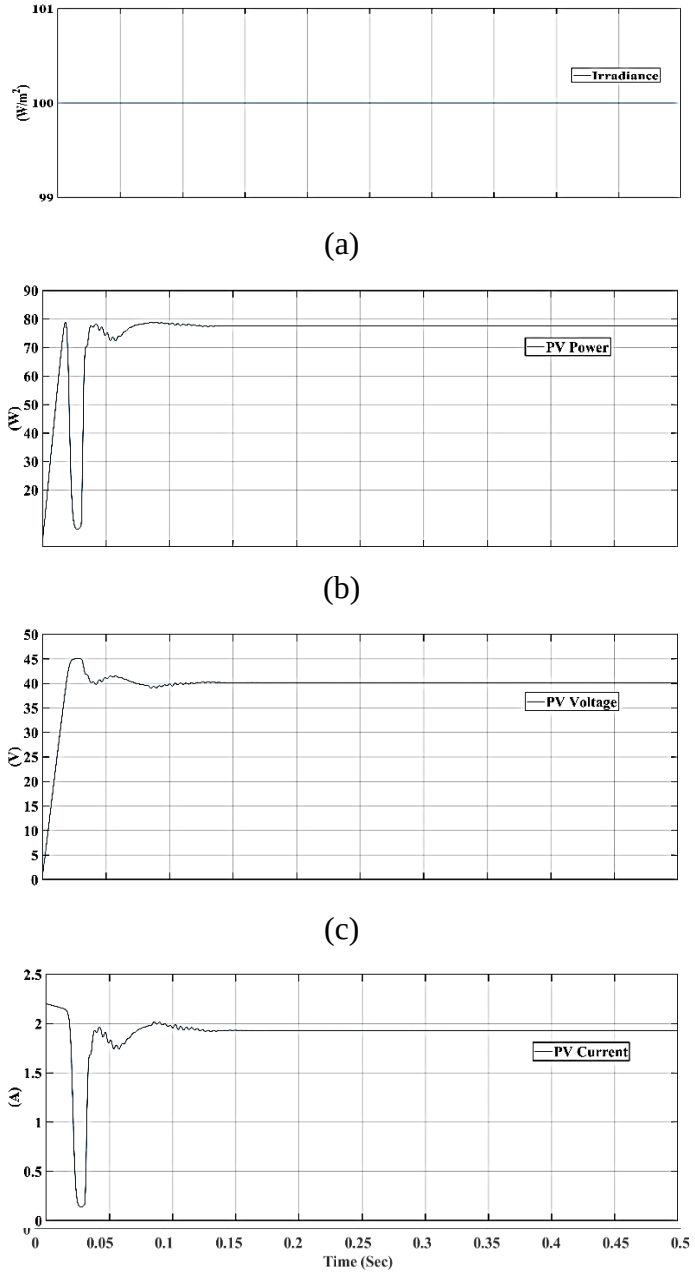


Fig 6.1: P-V characteristics of PV array at 100W/m^2

According to figure 6.2(a), which shows a constant irradiance of 100W/m^2 at which the simulation was performed, during this condition the MPP converter operating using DRL based control is able to obtain the maximum PV power of 76.5W shown in figure 6.2(b) with PV voltage and PV current of 40.26V and 1.9A shown by figure 6.2(c) & 6.2(d) respectively. Analyzing the obtained PV power and maximum available PV power, the efficiency of DRL based controller in tracking the PV power is 97.01% at 100W/m^2.



(a)

(b)

(c)

(d)

Fig 6.2: Performance of DRL agent under 100W/m^2 irradiance with a 500W DC load, (a)Constant irradiance of 100W/m2, (b) obtained PV power, (c)voltage across PV module, (d)current delivered by PV module

Since the load demand remains constant at 500W shown by figure 6.3(a) and the PV array is only able to supply 76.5W of solar power so, to keep the DC bus voltage constant the remaining power has to be supplied though the battery systems. According to figure 6.3(b), the battery system supply's the remaining load power of 429.5W which helps in maintaining a constant DC bus voltage of 48V as depicted by figure 6.3(c).

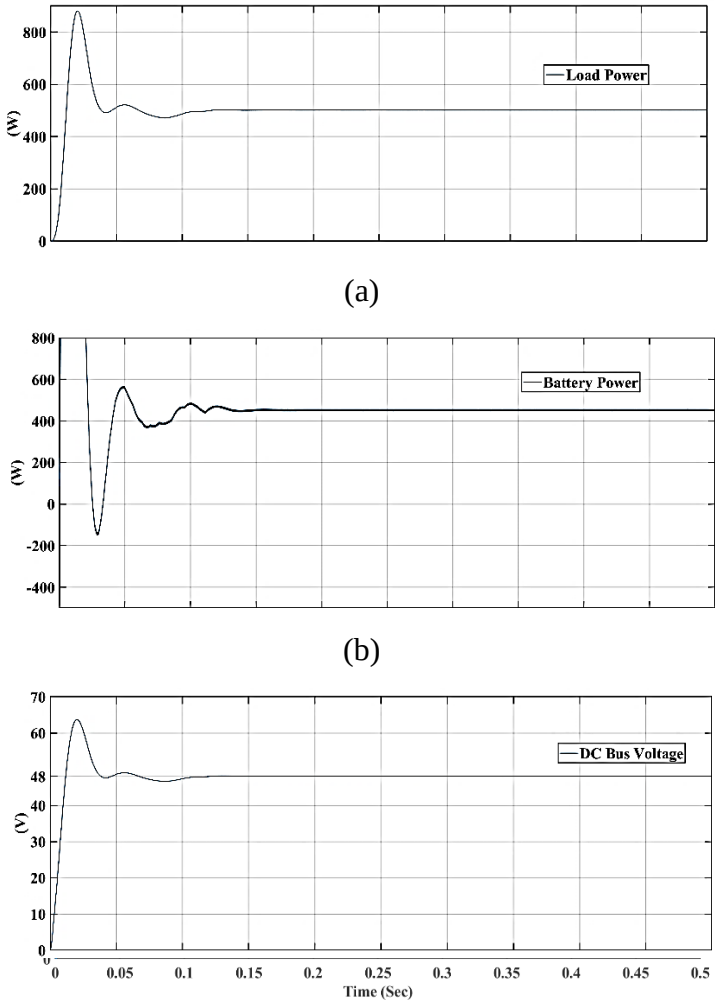


Fig 6.3: Performance of DRL agent under 100W/m^2 irradiance with a 500W DC load, (a) power demanded by the load, (b) power supplied/delivered to battery, (c) DC bus Voltage.

6.2. CASE 2

In this case, for analysis of deep reinforcement learning based solar battery system the PV array was operated at 600W/m^2 irradiance and according to figure 6.4, the maximum operating point power at this irradiance is 507.1W.

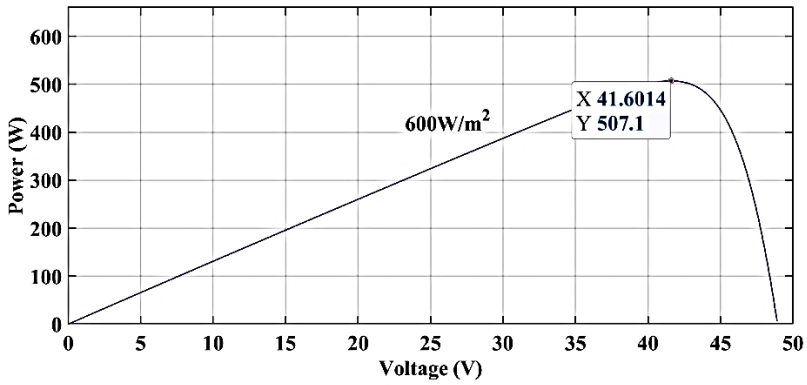
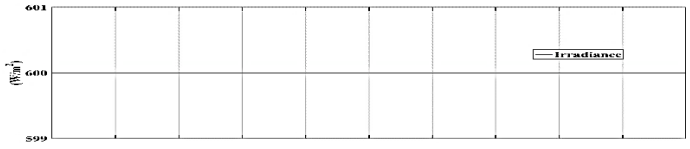
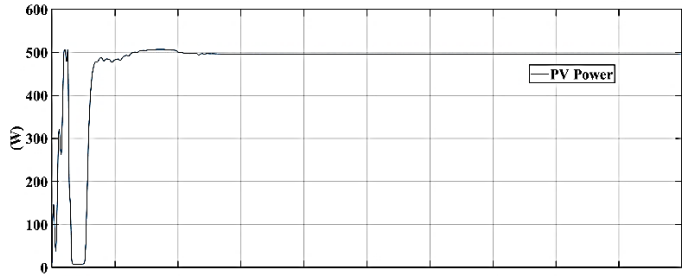


Fig 6.4: P-V characteristics of PV array at 600W/m^2

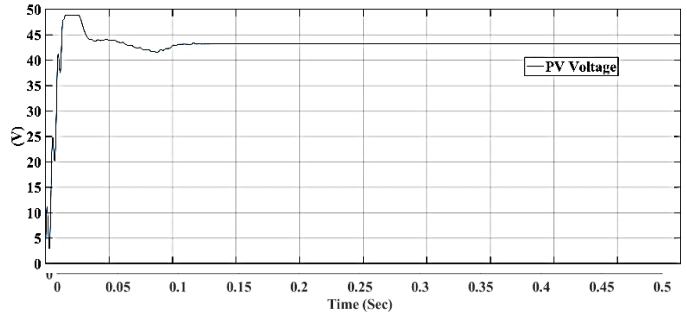
According to figure 6.5(a) which shows an irradiance level of 600W/m^2, the MPP converter is able to obtain 496.6W of PV power as shown in figure 6.5(b). The MPP converter operates the PV array with PV voltage and PV current as 42.6V and 11.65A respectively as shown by figures 6.5(c) & 6.5(d). Analysing the PV power obtained using the DRL based control and maximum available PV power, the converter operates with 97.92% efficiency at 600W/m^2.



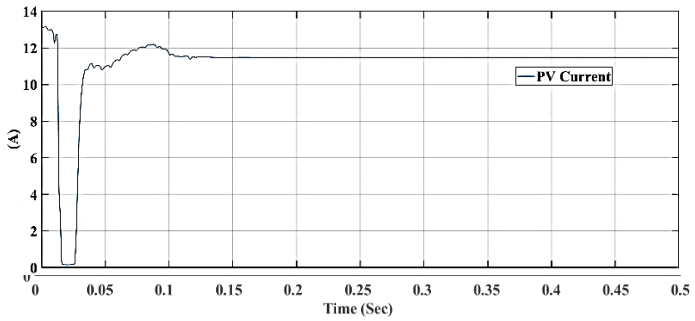
(a)



(b)



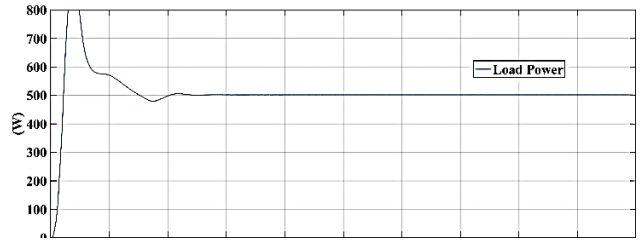
(c)



(d)

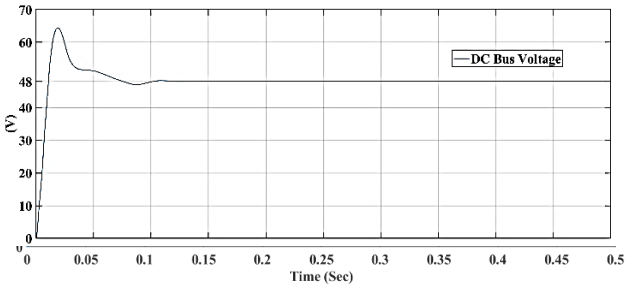
Fig 6.5: Performance of DRL agent under600W/m^{2}irradiance with a 500W DC load, (a)Constant irradiance of 600W/m^2, (b) obtained PV power, (c)voltage across PV module, (d)current delivered by PV module

As the load demand shown by figure 6.6(a) is constant at 500W, the solar power system is efficient in supplying the load power demand with minimal utilization of battery power at 6W as shown by figure 6.6(b). The battery power utilized in this scenario depicts the converter losses that the battery systems endure to maintain a constant DC bus voltage at 48V shown by figure 6.6(c)



(a)

(b)



(c)

Fig 6.6: Performance of DRL agent under 600W/m^{2}irradiance with a 500W DC load, (a) power demanded by the load, (b) DC bus Voltage, (c) power supplied/delivered to battery.

6.3. CASE 3

In this case 800W/m^2 irradiance was applied to PV array and according to figure 6.7 which represents the P-V characteristics of the PV array the maximum operating point power achievable is 680.15W.

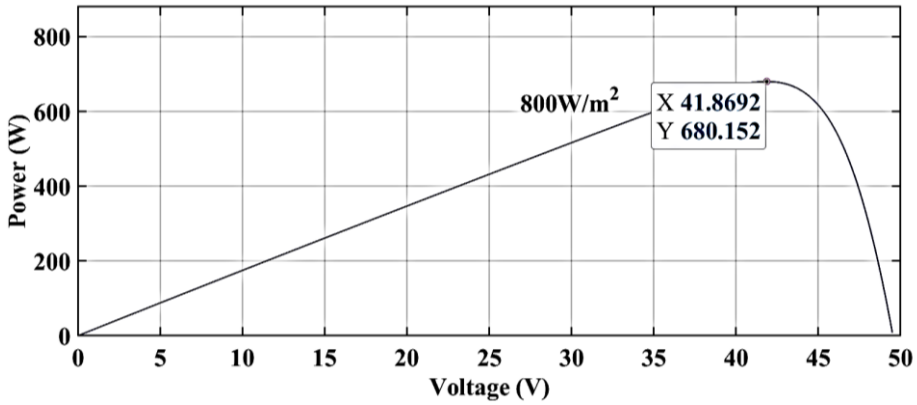
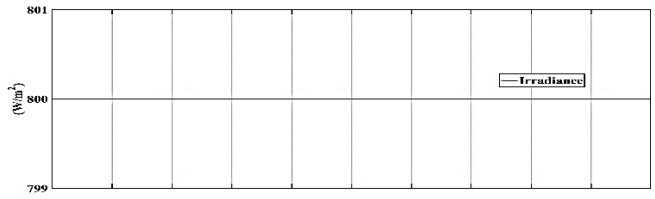
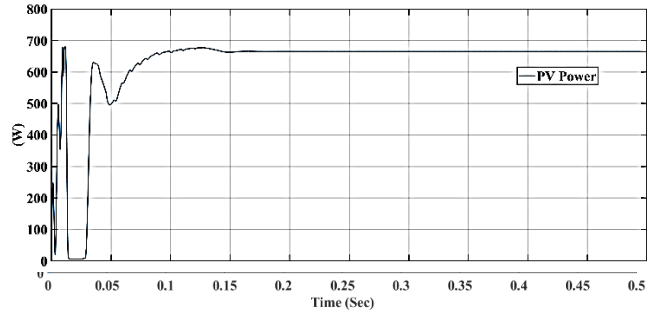


Fig 6.7: P-V characteristics of PV array at 800W/m^2

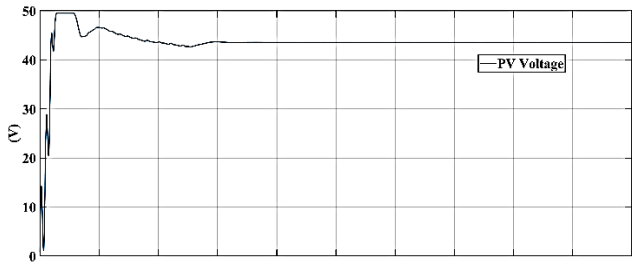
According to figure 6.8(a) which depicts an irradiance level of 800W/m^2, the MPP converter when operated at this irradiance level is able to operate the PV array to deliver 665.2W of PV power as shown by figure 6.8(b). This is achieved by keeping the PV voltage and PV current at 42.8V & 15.54A as observed by figures 6.8(c) & 6.8(d) respectively. This case reveals an efficiency of 97.80% in tracking the maximum PV power at 800W/m^2 using the DRL based control.



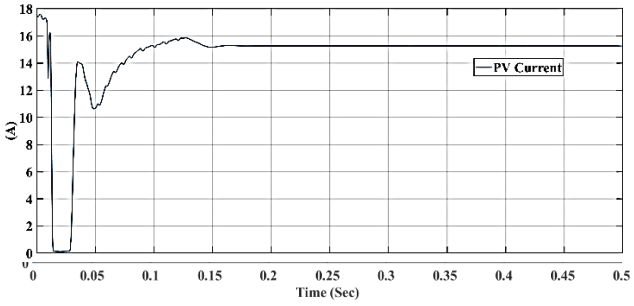
(a)



(b)



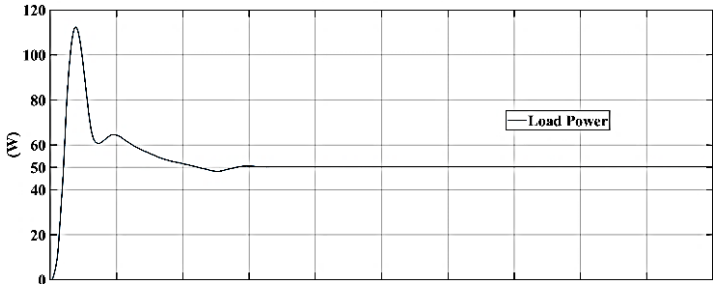
(c)



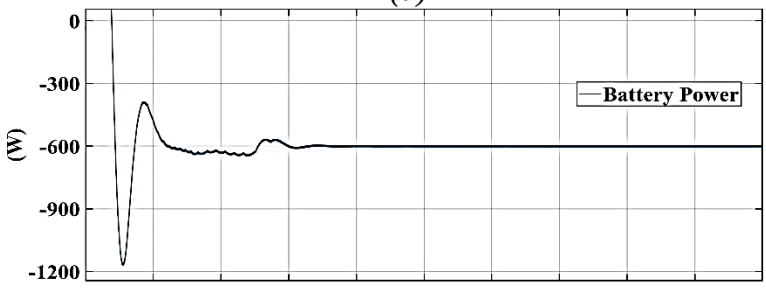
(d)

Fig 6.8: Performance of DRL agent under 800W/m^2 irradiance with a 50W DC load, (a)Constant irradiance of 800W/m^2, (b) obtained PV power, (c)voltage across PV module, (d)current delivered by PV module

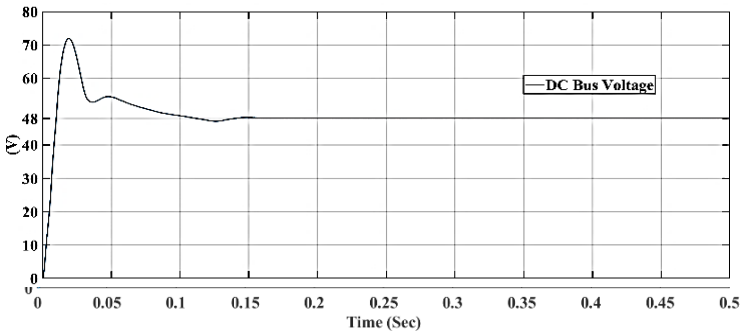
In this condition the load demand is kept low at 50W shown by figure 6.9(a), since the PV power is more than the load power demand thus the remaining PV power can be used in charging the battery. Observing figure 6.9(b) which shows that the battery is charging at 599.5W and during this operation the DC bus voltage is maintained constant at 48V shown by figure 6.9(c).



(a)



(b)



(c)

Fig 6.9: Performance of DRL agent under 800W/m^2 irradiance with a 50W DC load, (a) power demanded by the load, (b) power supplied/delivered to battery, (c) DC bus Voltage.

After observing the three operating cases, in which the PV array is operated at 100,600,800W/m^2 irradiance in case 1, 2, 3 respectively. In the first case, a 500W DC load is attached to the DC bus and since the PV array is operated at low irradiance the battery system had to be utilized in order to supply the load power for maintaining a constant DC bus voltage at 48V. In the second case with a medium irradiance level of 600W/m^2 the PV array is able to supply the load demand power of 500W with nominal power demanded from the battery system. In case 3, the load power demanded was 50W and as the PV array was able to operate at maximum power point condition delivering 665.2W of power, the remaining PV power was utilized in charging the battery system that can be used later to meet the load demands.

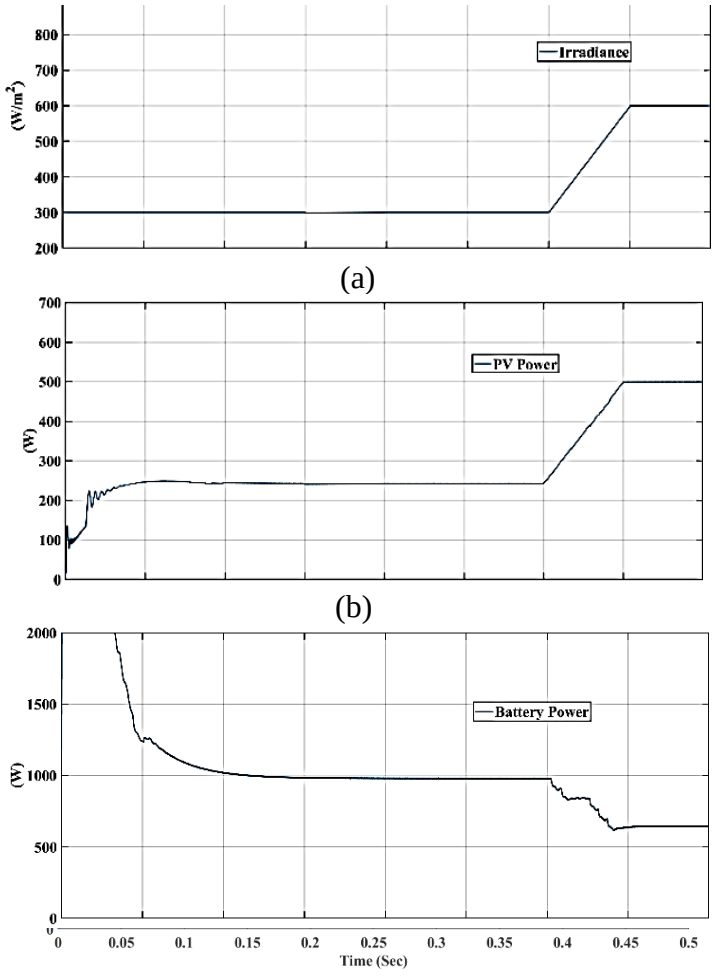
Cases 1 & 3, together depict the bidirectional operation of BDC in which the BDC operates as a buck converter while charging the battery and as a boost converter while discharging the battery. Observing these cases, the solar power management system works by operating the solar PV system to converter the solar power into electrical power which is then either utilized by the loads entirely or used to charge the battery system using the BDC converter.

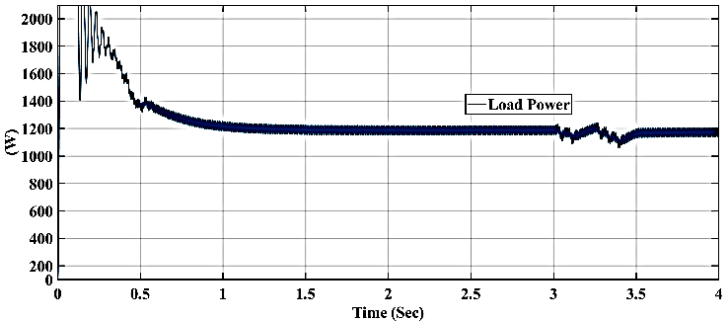
As the DRL based solar power management system is controlled by the deep reinforcement learning agent which observes the changes in the environment conditions by is observation states, as DC bus voltage and PV voltage were considered for making such observations so the agent reacts to any changes that occur in the observation states. Based upon these changes the neural networks associated with the actor network generates the actions which are the duty ratios for MPP converter and BDC converter, by maintaining a single control strategy for controlling the MPP converter and BDC

converter the DRL agent is able to quickly adjust to any changes and maintaining a constant DC bus voltage.

6.4. CASE 4

In this case, the PV array was operated with a variable irradiance that rises from 300W/m^2 to 600W/m^2 shown by figure 6.10(a) with a 1.2kW AC load which is connected to the inverter. According to figure 6.10(b), at irradiances of 300, 600W/m^2 the maximum PV power that the PV array is able to deliver is 243.3W and 498.6W respectively. According to figure 6.10(d) the load demand as observed at the DC bus is constant at 1.2kW thus the PV systems alone are inadequate in supplying the load demand hence, the battery systems have to be operated in for maintaining a constant DC bus voltage which can be observed by figure 6.10(c).

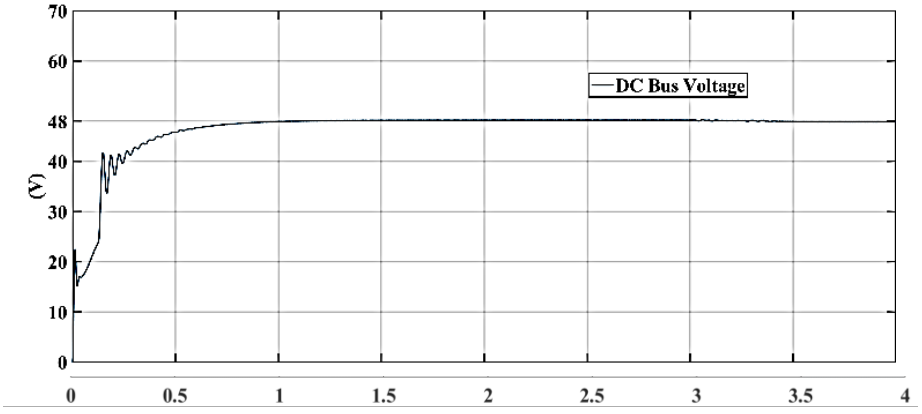




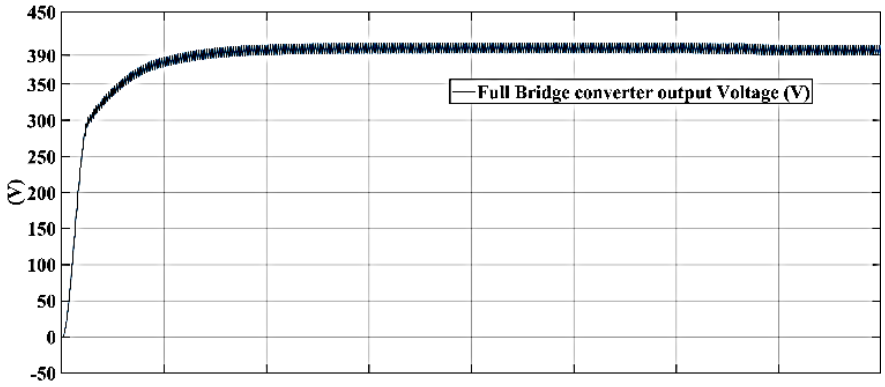
(d)

Fig 6.10: Performance of DRL agent at variable irradiance with constant 1.2kW AC load, (a) Variable irradiance levels applied to PV array, (b) PV power obtained at variable irradiance, (c) Battery power utilized, (d) Load power demand as observed at DC bus.

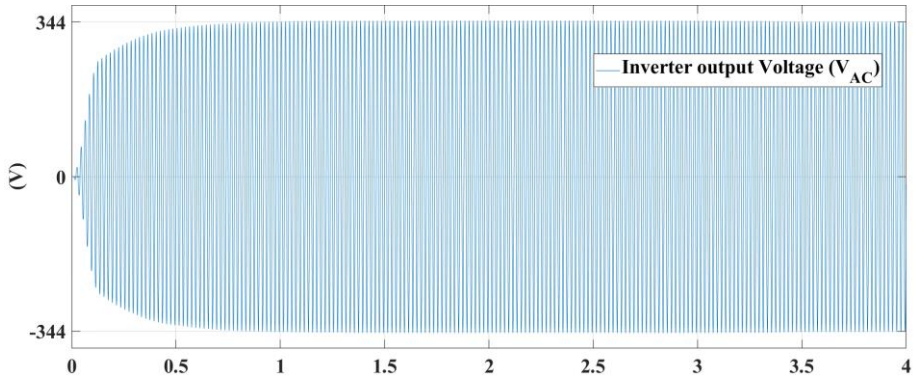
Observing the figures 6.11(b) & 6.11(c) the two-stage operation of inverter system can be seen, where first the bridge converter is operated to boost the 48V DC bus voltage shown by figure 6.11(a) to 390V DC as shown in figure 6.11(b) then later this boosted voltage is applied to a SPWM inverter that converters the DC power into AC power as shown in figure 6.11(d). Together the full bridge converter and SPWM inverter form the inverter system that is designed to supply power to the AC loads rated at 220-240V(rms) at 50Hz. According to Figure 6.12, the total harmonic distortion (THD) of the generated inverter output is 0.79% which represents the ratio of fundamental component to harmonics present in output AC waveform.



(a)



(b)



(c)

Fig 6.11: Two stage inverter operation, (a) 48V DC bus, (b) Full bridge converter output at 390V, (c) output AC voltage at 50Hz.

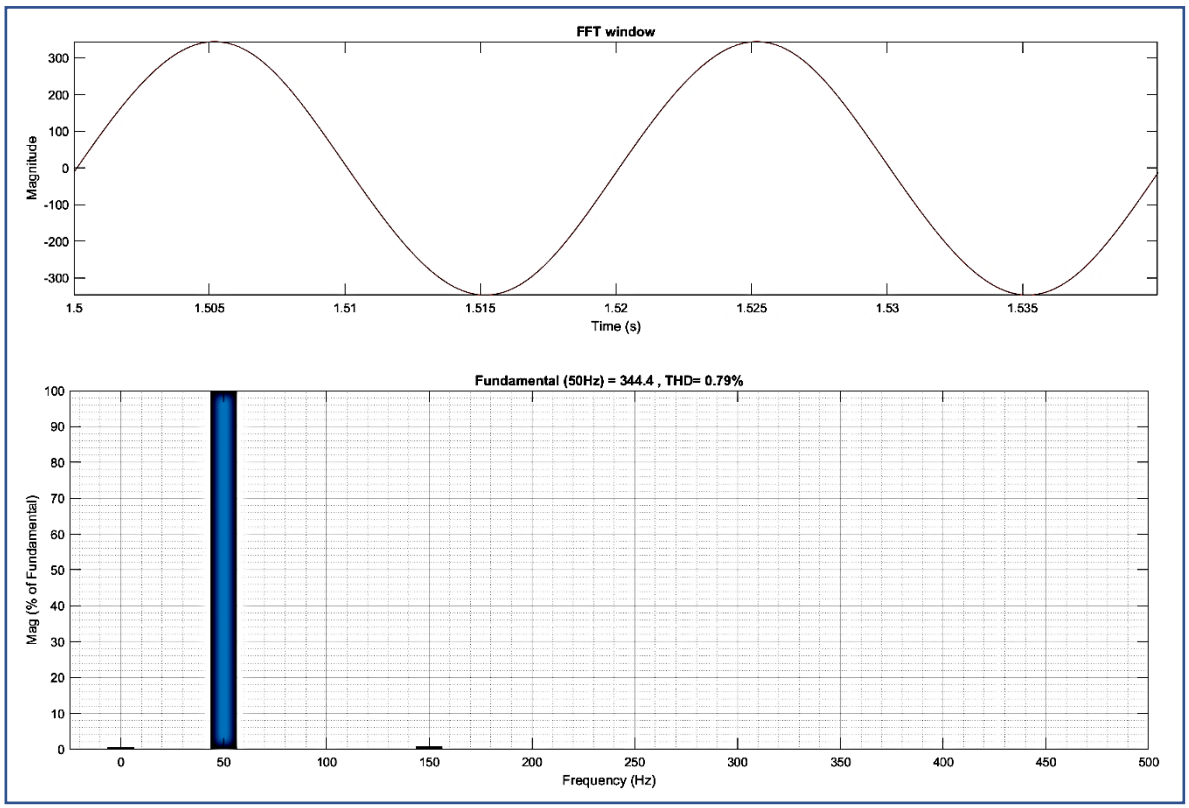
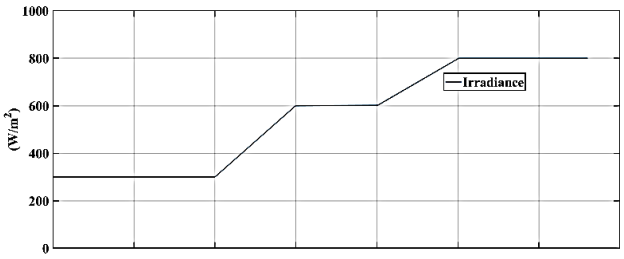


Fig 6.12: FFT analysis of AC voltage

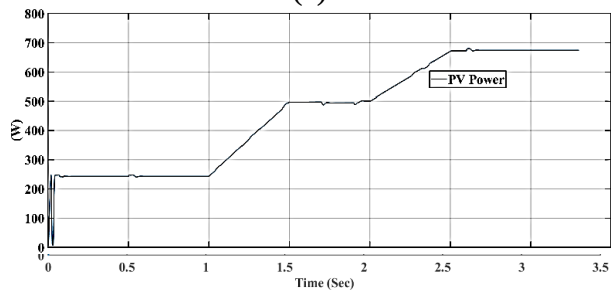
6.5. CASE 5

The simulation of variable loads at variable irradiation conditions is depicted by the figure 6.13, where according to figure 6.13(a) which depicts the variable irradiances that rise from 300 to 600 to 800W/m^2 that are applied to the PV array. At these irradiance levels the response of solar power management system is observed by changing the loads at different instances of time to see the changes in battery power and PV power.

According to figure 6.13(b) which shows the PV power, a rise in PV power with rise in irradiance can be observed. According to the P-V characteristics of PV array which also depicts a positive correlation between the irradiance and available solar power. Observing figure 6.13(c), the variable load demands that occur during the entire simulation depicts the simulated version of loads that can be connected to the DC bus throughout its operation. Since there may be situations where the load demand may be more than the available solar power then in those conditions, the battery discharge operations can be observed where the additional load power is provided by the battery. During conditions where load demand is lower than the available solar power then, the battery charge operations can be observed. Both the battery operating conditions can be seen by figure 6.13(d). By operating the PV array and BDC in conjugation to one another, the deep reinforcement learning based solar power management system is able to maintain a constant DC bus voltage shown in figure 6.13(e) while catering to the various load demands.



(a)



(b)

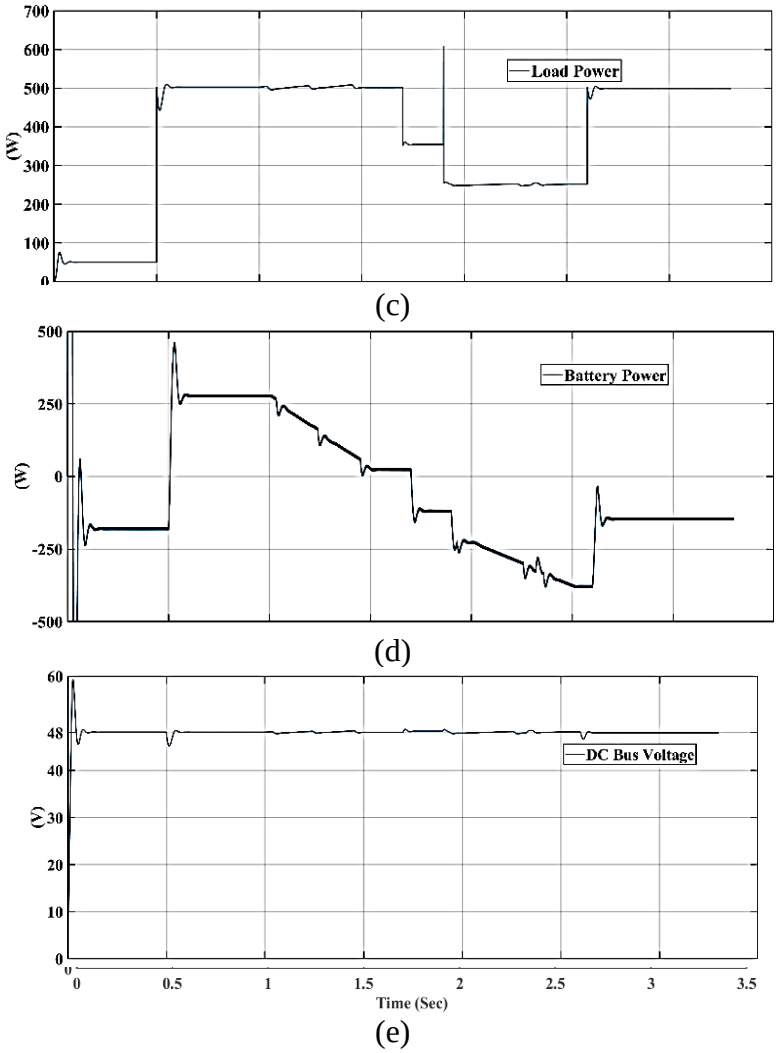


Fig 6.13: Operation of DRL based solar power management system under variable load and variable irradiance condition (a) variable irradiance applied (b) PV power obtained (c) Load power demand (d) Battery power utilization (e) DC bus voltage.

To analyse the efficiency of the deep reinforcement learning based solar power management system in tracking of maximum power point, figure 6.14 can be observed to obtain the ideal PV power available at a particular irradiance and compare it with the power obtained using DRL agent based MPPT control. The efficiency can be calculated as

$$\eta = \frac{P_{PV_{DRL}}}{P_{ideal}} \times 100 \qquad (6.1)$$

Where $P_{PV_{DRL}}$ is the PV power obtained using DRL agent and P_{ideal} is the ideal power that can be obtained from PV array.

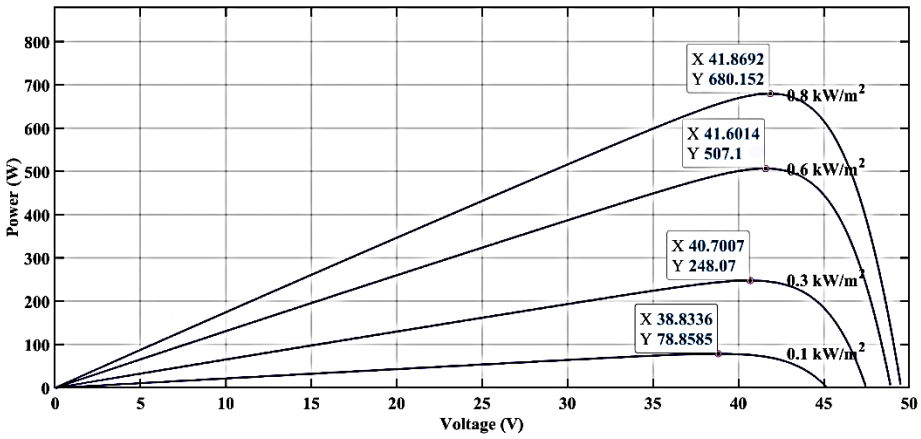


Fig 6.14: P-V characteristics of PV array at 100, 300, 600, 800W/m^2

According to table 6.2, that compares the power output of PV array using the DRL agent to the ideal power output at various irradiance levels, observations can be drawn regarding the efficiency of the deep reinforcement learning agent based maximum power point tracking technique.

Irradiance	**P_{ideal} (W)**	**$P_{PV_{DRL}}$ (W)**	**Efficiency (%)**
100	78.85	76.5	97.01
600	507.1	496.6	97.92
800	680.15	665.2	97.80

Table 6.2: comparison between the power obtained using DRL and ideal condition.

Based upon the observations made in the table 6.2, the DRL based maximum power point tracking method has an efficiency greater than 97% in chasing the maximum power point while simultaneously, operating the bidirectional converter in-order to maintain a constant DC bus voltage.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1. CONCLUSION

The proposed project shows a significant potential in development of a solar PV battery-based system which is controlled by deep reinforcement learning technique capable of attaining the maximum power point for a PV array in short duration of time. This method is an efficient MPPT method that is capable of operating at various load conditions and maintain a constant DC bus voltage. The benefit of this strategy is that it does not require any prior model of the control system. The controller learns how to respond in terms of action space based on the reward they receive through ongoing interactions that is represented as the observation space in the environment. Rather than employing a look-up table as in the case of RL-based technique, DRL employs neural networks to estimate a value function or a policy, allowing for a considerable reduction in the amount of memory required for action spaces.

The DRL algorithm is represented by the agent, and the action is the duty cycle perturbation. It begins by transmitting a prior state to the agent, that then responds to the previous state based on its knowledge. The environment then replies by returning a duo of the next state and a reward to the agent. Based on the reward and current condition acquired from the environment, the agent can learn how to respond. The suggested MPPT algorithms autonomously manages the changes of the duty cycle to find the optimal MPP after being trained on historical data gathered by direct contact with the solar power management system.

As the controller was designed using DDPG agent that enables the DRL based controller to continuously track changes in the operation conditions. This helps the DRL-agent to quickly adapt according to the changes and maintaining a constant DC bus voltage. The DRL agent tracks the changes in DC bus voltage and PV voltage which are defined in the observation space, these changes along with the reward generated by the previous state-action pair are then relayed to the DRL agent. The DRL agent uses a DDPG algorithm which is an actor-critic network-based algorithm that respond to these changes in state and the reward, determine the action which are the duty ratios for MPP converter and BDC. The MPP converter duty ratio operates the PV array at MPP whereas, the BDC duty ratio ensures that the charging and discharging of battery takes place in such a way that a constant DC bus voltage can be maintained. By utilising this DRL based solar power management system efficiency greater than 97% was observed in tracking solar power while maintaining a constant DC bus voltage for the attached loads.

7.2. FUTURE SCOPE

For the scope of future development of this project, MPPT algorithm can be built using rule-based reinforcement learning (RuRL) which combines reinforcement learning and traditional knowledge-based techniques. Where to address environmental navigation challenges, rule-based approaches rely on the model of environment and in-depth information of the system, and tends to converge early to suboptimal policies. This is achieved by reducing redundant explorations and guiding the explorations to obtain faster convergence.

Using artificial intelligence more accurate weather forecasting models can developed for irradiation prediction. The weather forecast models can be incorporated into the DRL agent for better scheduling of PV and battery power. Using various deep learning techniques such as, long short-term memory (LSTM) and generative adversarial network (GAN) can be used as control strategy in inverter system and mitigating AC harmonics that can be computed by analysing the voltage and current behaviours at the AC load side.

APPENDIX 1

ARTIFICIAL NEURAL NETWORKS

The human brain's typical functioning inspired the invention of ANNs, but they are only distantly linked to their organic counterparts. Although ANNs do not reach the intricacy of the brain, biological neural networks and ANNs have two crucial characteristics. To begin, both networks' building blocks are simple computational devices that are closely linked. Second, the network's function is determined by the connections between neurons. ANNs are computing networks that are parallel distributed which behave similarly to neurons in the brain. There are multiple input signals [$X = (x_1, x_2, ., x_n)$] to neurons in some basic properties shown by figure A 1.1. Each input is given a relative weight [$W = (w_1, w_2, ., w_n)$] that influences its impact. This is similar to biological neurons' varied synaptic strengths. In the way they interact to form an impulse, some inputs are more essential than others. The intensity of the input signal is justified by weights, which are modifiable coefficients within the network. The summation block, which correlates to the human cell, produces the neuron's output signal as summation of all the input weights.

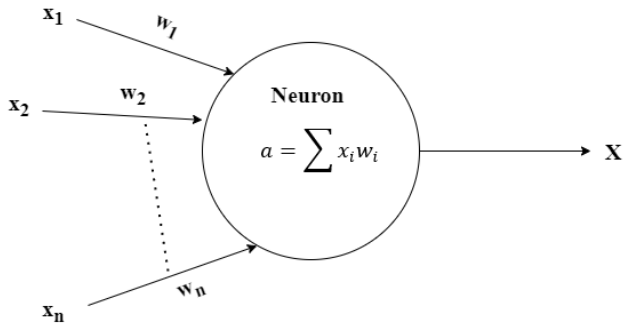


Fig A 1.1: Schematic diagram of artificial neuron as basis processing unit.

A1.1. STRUCTURE OF MULTILAYER NETWORK

The three layers (input, hidden, and output) of nonlinear computational elements make up a neural network. Through the hidden layer, data moves from the input layer to the output layer shown by figure A 1.2.

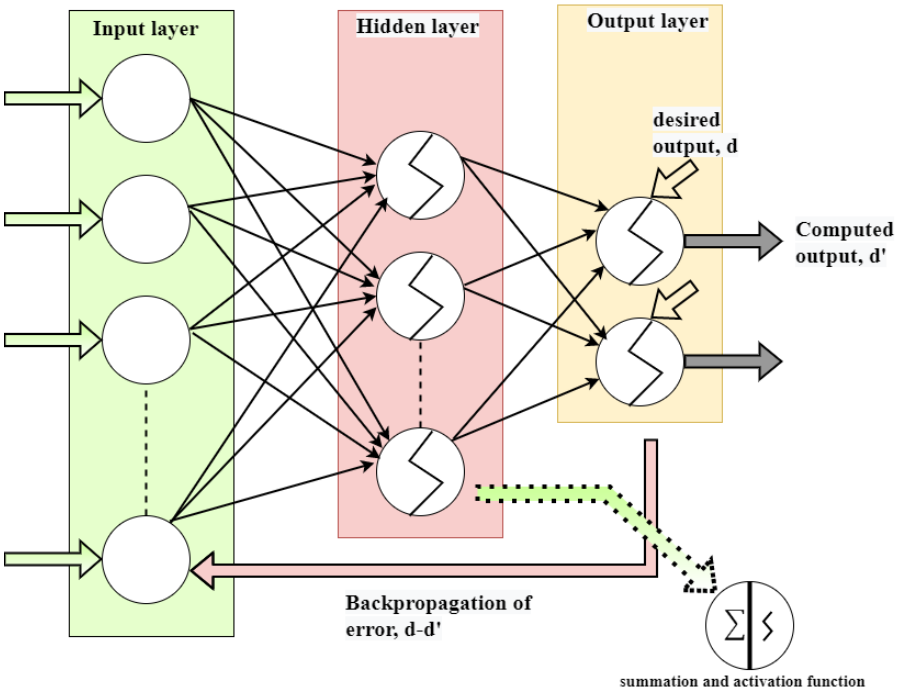


Fig A 1.2: Structure of multilayer network

The neurons in each layer are completely interconnected with those in the layers above and below it. These connections are represented as weights in the computational process. The weights play an important part in the signal's transmission over the network. They are used to store the knowledge of the problem-solution connection held by the neural network. The input layer's neurons are proportional to the number of independent variables in the model, but the output layer's neurons are fixed. It is possible to utilise a single or multiple number of output neurons. Furthermore, the number of hidden layers and their neurons are proportional to the model's complexity and are important parameters in the evolution of the NN model.

A neural network is trained to minimise the differences between the anticipated goal values and the model's output values. The weights are changed to decrease errors if the network returns the incorrect response or if the errors exceed a certain threshold. As a result, errors are reduced, and subsequent network responses are more likely to be correct. The network is provided with the inputs and anticipated targets

serially during the learning procedure. The learning method includes a forward-propagation step tailed by a stride of backward propagation.

A1.2. FORWARD-PROPAGATION

The forward-propagation starts with the input layer receiving an input pattern. The network collects information from input and output neurons, then summarises it. To reduce the error between the anticipated target and the model's output, the training is an iterative gradient technique. Assuming a neuron j of hidden layer p is used to create a network z with input layer represented as x_{ji} and weights w with a sigmoid activation function σ, the network output can be represented as;

$$\emptyset(z) = \sigma(z) = \frac{1}{1+e^{-z}} \quad \text{(A 1.1)}$$

The differential of ∅(z) can be represented as

$$\emptyset'(z) = \emptyset(z)\big(1 - \emptyset(z)\big) \quad \text{(A 1.2)}$$

Assuming the output error E is computed as mean square error against the desired output y_d,

$$E = \frac{1}{2}\sum_{d\in D}\big(y_d - \sigma(z)\big)^2 \quad \text{(A 1.3)}$$

Since a multilayer network consists of multiple neurons form input layers connected to a single layer hence, the gradient of error is calculated against the weights w_i,

$$\frac{dE}{dw_i} = \frac{1}{2}\sum_d \frac{dE}{d\hat{y}_d}.\frac{d\hat{y}_d}{dw_i} \quad \text{(A 1.4)}$$

Where $\hat{y}_d = y_d - \sigma(z) = y_d - \sigma(w.X_d)$ and represents the networks output,

$$\frac{dE}{dw_i} = \sum_d (y_d - \hat{y}_d).\frac{d(y_d - \sigma(z))}{dw_i} \quad \text{(A 1.5)}$$

$$\frac{dE}{dw_i} = \sum_d (y_d - \hat{y}_d).x_{id}.\hat{y}_d(1 - \hat{y}_d) \quad \text{(A 1.6)}$$

A1.3. BACK PROPAGATION

The error value assigned to each neuron represents the degree of mistake connected with that neuron. As a result, the neuron backpropagates the appropriate weight adjustment, and the weights of the output neurons are updated. The backpropagation of error is calculated by the following algorithm where y_j represents the output of network j

with activation function O_j and summation function net_j. The error associated with the summation component is represented by $\delta_j = \frac{dE}{dO_j} . \frac{dO_j}{dnet_j}$.

Set all weights as small random numbers

Until completed **do,**

For each training episode **do,**

Input training example into network and calculate output

For each output unit k,

$$\delta_k \leftarrow O_k(1 - O_k)(y_k - O_k)$$

For each hidden unit h,

$$\delta_h \leftarrow O_h(1 - O_h) \sum_{k \in output} w_{h,k} \delta_k$$

Update each weight w_{ij}

$$w_{ij} \leftarrow w_{i,j} + \Delta w_{i,j}$$

$$where\ \Delta w_{i,j} = \eta \delta_j x_{i,j}$$

A.1.4. ACTIVATION FUNCTIONS

An Activation Function governs whether or not a neuron is to be triggered. This is done using simple mathematical operations which control whether the neuron's input is important or not. The Activation Function's generates output from a collection of input values that are fed into a node (or a layer).

Some most commonly used activation functions are

a. Sigmoid

As represented by figure A 1.3(a), the sigmoid function can be expressed as

$$f(x) = \frac{1}{1+e^{-x}} \quad \text{(A 1.7)}$$

b. Tanh

As represented by figure A 1.3(b), the tanh function can be expressed as

$$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad \text{(A 1.8)}$$

c. Rectified Linear Unit (ReLU)

As represented by figure A 1.3(c), the ReLU function can be expressed as

$$f(x) = \max(0, x) \quad \text{(A 1.9)}$$

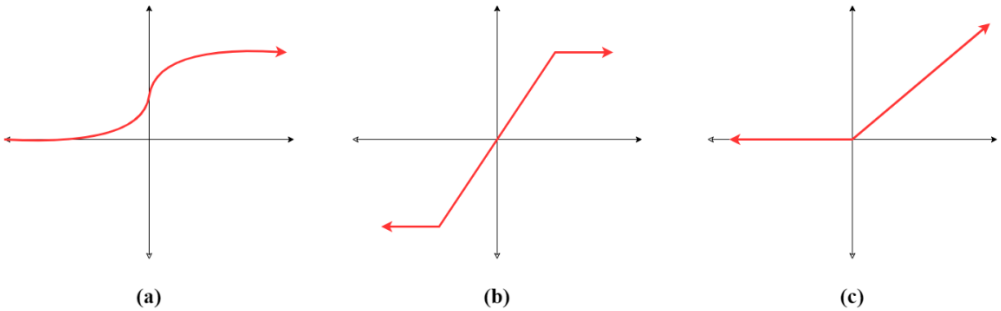


Fig A 1.3: Activation functions used in neural networks (a) Sigmoid function (b) tanh function (c) ReLU function

REFERENCES

[1.] Faiza Belhachat, Cherif Larbes, “A review of global maximum power point tracking techniques of photovoltaic system under partial shading conditions” , *Renewable and Sustainable Energy Reviews* , Volume 92, 2018, Pages 513-553, ISSN 1364-0321.

[2.] Phan, Bao & Lai, Ying-Chih & Lin, Chin. (2020). “A Deep Reinforcement Learning-Based MPPT Control for PV Systems under Partial Shading Condition.” *Sensors*. 20. 3039. 10.3390/s20113039.

[3.] F. Grouz and L. Sbita, "A safe and easy methodology for design and sizing of a stand-alone hybrid PV-wind system," *2014 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, 2014, pp. 1-8, doi: 10.1109/CISTEM.2014.7077043.

[4.] Li, Jing & Wei, Wei & Xiang, Ji. (2012). “A Simple Sizing Algorithm for Stand-Alone PV/Wind/Battery Hybrid Microgrids”. *Energies. 5.* 5307-5323. 10.3390/en5125307.

[5.] R. Hill, "Grid connected solar PV," *IEE Colloquium on Developments in Photovoltaic Electricity Production (Digest No: 1997/069)*, 1997, pp. 4/1-4/3, doi: 10.1049/ic:19970366.

[6.] F. Blaabjerg, R. Teodorescu, M. Liserre and A. V. Timbus, "Overview of Control and Grid Synchronization for Distributed Power Generation Systems," in *IEEE Transactions on Industrial Electronics, vol. 53*, no. 5, pp. 1398-1409, Oct. 2006, doi: 10.1109/TIE.2006.881997.

[7.] Teodorescu, R.; Blaabjerg, F.; Liserre, M.; Loh, P.C.: “Proportional-resonant controllers and filters for grid-connected voltage-source converters”, *IEE Proceedings - Electric Power Applications*, 2006, 153, (5), p. 750-762, DOI: 10.1049/ip-epa:20060008

[8.] Y. Libin et al., "A New Theory of Reactive Power Control of Grid Connected PV Inverter," *2015 International Conference on Intelligent Transportation, Big Data and Smart City*, 2015, pp. 35-38, doi: 10.1109/ICITBS.2015.15.

[9.] M.R. Khan, G.Mul. “The Integration and Control of Multifunctional Stationary PV-Battery Systems in Smart Distribution Grid [C]”. *28th EU PVSEC*, Paris, 2013.

[10.] Z. Wang, X. Li, G. Li. “Energy storage control for the PV generation system in micro-grid [C]”. *5th Int. Conf. Critical Infrastructure*, Sep. 2010: 1-5

[11.] B. Liu, Y. Zhang, B. Hu, L. Xue and Z. Wei, "An integrated control strategy of PV-battery hybrid systems," *2014 International Power Electronics and Application Conference and Exposition*, 2014, pp. 419-422, doi: 10.1109/PEAC.2014.7037892.

[12.] J. Tant, F. Geth, D. Six, P. Tant and J. Driesen, "Multiobjective battery storage to improve PV integration in residential distribution grids," *2013 IEEE Power & Energy Society General Meeting*, 2013, pp. 1-1, doi: 10.1109/PESMG.2013.6672106.

[13.] X. Vallve, A. Graillot, S. Gual, et al. “Micro storage and Demand Side Management in distributed PV grid-connected installations [C]”. *9th Electrical Power Quality and Utilization international conference*, Oct. 2007.

[14.] M. V. Nair and Lekshmi S., "Reconfigurable Solar Converter for PV battery application," *2014 International Conference on Electronics, Communication and Computational Engineering (ICECCE)*, 2014, pp. 61-66, doi: 10.1109/ICECCE.2014.7086636.

[15.] K. Sun, L. Zhang, Y. Xing and J. M. Guerrero, "A Distributed Control Strategy Based on DC Bus Signaling for Modular Photovoltaic Generation Systems With Battery Energy Storage," *in IEEE Transactions on Power Electronics, vol. 26*, no. 10, pp. 3032-3045, Oct. 2011, doi: 10.1109/TPEL.2011.2127488.

[16.] "IEEE Guide for Array and Battery Sizing in Stand-Alone Photovoltaic (PV) Systems," in *IEEE Std 1562-2007* , vol., no., pp.1-32, 12 May 2008, doi: 10.1109/IEEESTD.2008.4518937.

[17.] A. Nigam and A. Kumar Gupta, "Performance and simulation between conventional and improved perturb & observe MPPT algorithm for solar PV cell using MATLAB/Simulink," *2016 International Conference on Control, Computing, Communication and Materials (ICCCCM), 2016*, pp. 1-4, doi: 10.1109/ICCCCM.2016.7918220.

[18.] J. Lee, J. Ko, D. Jeong and D. Chung, "MPPT control of photovoltaic system using the two-mode IC method," 2012 12th International Conference on Control, Automation and Systems, 2012, pp. 1135-1138.

[19.] Kofinas, P. & Doltsinis, Stefanos & Dounis, Anastasios & Vouros, George. (2017). “A Reinforcement Learning Approach for MPPT Control Method of Photovoltaic Sources.” *Renewable Energy. 108. 461-473. 10.1016/j.renene.2017.03.008*.

[20.] Roy Chaoming Hsu, Cheng-Ting Liu, Wen-Yen Chen, Hung-I Hsieh, Hao-Li Wang, "A Reinforcement Learning-Based Maximum Power Point Tracking Method for Photovoltaic Array", *International Journal of Photoenergy*, vol. 2015, Article ID 496401, 12 pages, 2015.

[21.] R. Hussan and A. Sarwar, "Maximum Power Point Tracking Techniques under Partial Shading Condition- A Review," *2018 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 2018, pp. 293-298, doi: 10.1109/ICPEICES.2018.8897493.

[22.] Lillicrap, Timothy & Hunt, Jonathan & Pritzel, Alexander & Heess, Nicolas & Erez, Tom & Tassa, Yuval & Silver, David & Wierstra, Daan. (2015). “Continuous control with deep reinforcement learning”. *corr*.

[23.] L. Avila, M. De Paula, I. Carlucho and C. Sanchez Reinoso, "MPPT for PV systems using deep reinforcement learning algorithms," in *IEEE Latin America Transactions, vol. 17*, no. 12, pp. 2020-2027, December 2019, doi: 10.1109/TLA.2019.9011547.

[24.] T. Rout, M. K. Maharana, A. Chowdhury and S. Samal, "A Comparative study of Stand-alone Photo-Voltaic System with Battery storage system and Battery Supercapacitor storage system," *2018 4th International Conference on Electrical Energy Systems (ICEES),* 2018, pp. 77-81, doi: 10.1109/ICEES.2018.8442346.

[25.] Cota, A., Ghassemi, M., & Foster, R. (2009). “Solar Energy: Renewable Energy and the Environment”. *CRC Press*.

[26.] C. E. Lin and B. C. Phan, "Optimal Hybrid Energy Solution for Island Micro-Grid," 2016 *IEEE International Conferences on Big Data and Cloud Computing (BDCloud),* Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), 2016, pp. 461-468, doi: 10.1109/BDCloud-SocialCom-SustainCom.2016.74.

[27.] Alivarani Mohapatra, Byamakesh Nayak, Priti Das, Kanungo Barada Mohanty, “A review on MPPT techniques of PV system under partial shading condition”, *Renewable and Sustainable Energy Reviews*, Volume 80, 2017, Pages 854-867, ISSN 1364-0321

[28.] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. A. Cañizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke, G. A. Jiménez-Estévez, and N. D. Hatziargyriou, “Trends in microgrid control*,” IEEE Trans.Smart Grid, vol. 5*, no. 4, pp. 19051919, Jul. 2014.

[29.] N. Bizon, N. Mahdavi Tabatabaei, and H. Shayeghi, “Analysis, Control and Optimal Operations in Hybrid Power Systems”. London, U.K.: *Springer*,2014.

[30.] Luu Ngoc An and Tran Quoc-Tuan, "Optimal energy management for grid connected microgrid by using dynamic programming method,*" 2015 IEEE Power & Energy Society General Meeting*, 2015, pp. 1-5, doi: 10.1109/PESGM.2015.7286094.

[31.] G. Amar, B. EL-Madjid and T. Abdelaziz, "Optimal control and energy management of grid-connected PV-diesel-battery hybrid power system," *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, 2019, pp. 1-6, doi: 10.1109/ICAEE47123.2019.9014656.

[32.] Y. Riffonneau, S. Bacha, F. Barruel, and S. Ploix, ``Optimal power low management for grid connected PV systems with batteries," *IEEE Trans. Sustain. Energy, vol. 2, no. 3*, pp. 309320, Jul. 2011.

[33.] L. Ngoc An and T. Quoc-Tuan, ``Optimal energy management for grid connected microgrid by using dynamic programming method,*" in Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 15.

[34.] D. Zhang, ``Optimal design and planning of energy microgrids," *Ph.D. dissertation, Dept. Chem. Eng*., Univ. College London, London, U.K.,

[35.] Yousra Shaiek, Mouna Ben Smida, Anis Sakly, Mohamed Faouzi Mimouni, “Comparison between conventional methods and GA approach for maximum power point tracking of shaded solar PV generators”, *Solar Energy*, Volume 90, 2013, Pages 107-122, ISSN 0038-092X.

[36.] Ahmed, J.; Salam, Z. “A Maximum Power Point Tracking (MPPT) for PV system using Cuckoo Search with partial shading capability”. *Appl. Energy* 2014, 119, 118–130.

[37.] R. B. Bollipo, S. Mikkili and P. K. Bonthagorla, "Hybrid, optimal, intelligent and classical PV MPPT techniques: A review," *in CSEE Journal of Power and Energy Systems, vol. 7, no. 1*, pp. 9-33, Jan. 2021, doi: 10.17775/CSEEJPES.2019.02720.

[38.] G. Willcox and L. Rosenberg, "Short Paper: Swarm Intelligence Amplifies the IQ of Collaborating Teams," *2019 Second International Conference on Artificial Intelligence for Industries (AI4I)*, 2019, pp. 111-114, doi: 10.1109/AI4I46381.2019.00036.

[39.] M. Hijjo, F. Felgner, and G. Frey, ``PV-Battery-Diesel microgrid layout design based on stochastic optimization," *in Proc. 6th Int. Conf. Clean Electr. Power (ICCEP), Jun. 2017*, pp. 30-35.

[40.] U. B. Tayab, F. Yang, M. El-Hendawi, and J. Lu, ``Energy management system for a grid-connected microgrid with photovoltaic and battery energy storage system," *in Proc. Austral. New Zealand Control Conf. (ANZCC)*, Dec. 2018, pp. 141-144.

[41.] A. T. Eseye, D. Zheng, J. Zhang, and D. Wei, ``Optimal energy management strategy for an isolated industrial microgrid using a modified particle swarm optimization,*'' in Proc. IEEE Int. Conf. Power Renew. Energy (ICPRE)*, Oct. 2016, pp. 494-498.

[42.] M. Kirschenbaum and D. W. Palmer, "Perceptualization of particle swarm optimization,*" 2015 Swarm/Human Blended Intelligence Workshop (SHBI)*, 2015, pp. 1-5, doi: 10.1109/SHBI.2015.7321681.

[43.] P. K. Bharne, V. S. Gulhane and S. K. Yewale, "Data clustering algorithms based on Swarm Intelligence," *2011 3rd International Conference on Electronics Computer Technology*,2011, pp.407-411, doi: 10.1109/ICECTECH.2011.5941931.

[44.] D. Modani and V. Shrivastava, "Improvement in efficiency of PV module using soft computing based MPPT," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 2273-2278.

[45.] A. O. Erick and K. A. Folly, “Reinforcement learning approaches to power management in grid-tied microgrids: A review,” *in Proc. Clemson Univ. Power Syst. Conf.*, 2020, pp. 1-6.

[46.] R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction”, *2nd ed. Cambridge, MA, USA: MIT Press*, 2018.

[47.] Jiang, L.L.; Srivatsan, R.; Maskell, D.L. “Computational intelligence techniques for maximum power point tracking in PV systems: A review”. *Renewable Sustainable Energy* Rev. *2018*, 85, 14–45.

[48.] Glavic, M. “(Deep) Reinforcement learning for electric power system control and related problems: A short review and perspectives”. *Annu. Rev. Control 2019*, 48, 22–35.

[49.] Kahn, G.; A Villaflor, B.D.; Abbeel, P.; Levine, S. “Self-Supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation”, In Proceedings of *the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 20–25 May 2018; pp. 1–8.

[50.] X. Lu, X. Xiao, L. Xiao, C. Dai, M. Peng and H. V. Poor, "Reinforcement Learning-Based Microgrid Energy Trading With a Reduced Power Plant Schedule," *in IEEE Internet of Things Journal, vol. 6, no. 6*, pp. 10728-10737, Dec. 2019, doi: 10.1109/JIOT.2019.2941498.

[51.] E. O. Arwa and K. A. Folly, "Reinforcement Learning Techniques for Optimal Power Control in Grid-Connected Microgrids: A Comprehensive Review,*" in IEEE Access, vol. 8, pp. 208992-209007, 2020*, doi: 10.1109/ACCESS.2020.3038735.

[52.] Y.-S. Park, S. Lek, “Chapter 7 - Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling”, Editor(s): Sven Erik Jørgensen, *Developments in*

Environmental Modelling, Elsevier, Volume 28, 2016, Pages 123-140, ISSN 0167-8892, ISBN 9780444636232

[53.] Wang Zhifu, Wang Yupu, Rong Yinan, “Design of closed-loop control system for a bidirectional full bridge DC/DC converter”, *Applied Energy, Volume 194, 2017, Pages 617-625,* ISSN 0306-2619

[54.] Avila, Luis, Mariano De Paula, Maximiliano Trimboli, and Ignacio Carlucho. "Deep reinforcement learning approach for MPPT control of partially shaded PV systems in Smart Grids*." Applied Soft Computing* 97 (2020): 106711.

[55.] Bingöl, Okan, and Burçin ÖZKAYA. "A comprehensive overview of soft computing based MPPT techniques for partial shading conditions in PV systems." *Mühendislik Bilimleri ve Tasarım Dergisi 7*, no. 4 (2019): 926-939.

[56.] Singh, Yaduvir, and Nitai Pal. "Reinforcement learning with fuzzified reward approach for MPPT control of PV systems." *Sustainable Energy Technologies and Assessments* 48 (2021): 101665.

[57.] Chou, Kuan-Yu, Chia-Shiou Yang, and Yon-Ping Chen. "REINFORCEMENT LEARNING BASED MAXIMUM POWER POINT TRACKING CONTROL OF PARTIALLY SHADED PHOTOVOLTAIC SYSTEM." *Journal of Marine Science and Technology* 28, no. 5 (2020): 13.

[58.] Chou, Kuan-Yu, Chia-Shiou Yang, and Yon-Ping Chen. "Deep Q-Network Based Global Maximum Power Point Tracking for Partially Shaded PV System." In *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, pp. 1-2. IEEE, 2020.

[59.] Khan, Mohammad Junaid, and Lini Mathew. "Artificial intelligence based maximum power point tracking algorithm for photo-voltaic system under variable environmental conditions." In *2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE)*, pp. 114-119. IEEE, 2017.

[60.] Kermadi, Mostefa, and El Madjid Berkouk. "Artificial intelligence-based maximum power point tracking controllers for Photovoltaic systems: Comparative study." *Renewable and Sustainable Energy Reviews* 69 (2017): 369-386.

[61.] Ram, J. Prasanth, T. Sudhakar Babu, and N. Rajasekar. "A comprehensive review on solar PV maximum power point tracking techniques." *Renewable and Sustainable Energy Reviews* 67 (2017): 826-847.

[62.] Jiang, Lian L., R. Srivatsan, and Douglas L. Maskell. "Computational intelligence techniques for maximum power point tracking in PV systems: A review." *Renewable and Sustainable Energy Reviews* 85 (2018): 14-45.