Article

Deep Reinforcement Learning for Stability Enhancement of a Variable Wind Speed DFIG System†

Rahul Kosuru, Shichao Liu\* (Senior Member, IEEE), Wei Shi

|  |
| --- |
| **Citation:** Kosuru, R. and Liu, S, *Actuators* **2022**, *11*, x. https://doi.org/10.3390/xxxxx  Academic Editor: Firstname Lastname  Received: date  Accepted: date  Published: date  **Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.    **Copyright:** © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). |

Carleton University, Ottawa, Ontario, Canada ; [rahulkosuru@cmail.carleton.ca](mailto:rahulkosuru@cmail.carleton.ca); [shichaoliu@cunet.carleton.ca](mailto:shichaoliu@cunet.carleton.ca); [weishi@cunet.carleton.ca](mailto:weishi@cunet.carleton.ca)

**\*** Correspondence: [shichaoliu@cunet.carleton.ca](mailto:shichaoliu@cunet.carleton.ca)

**Author to whom correspondence should be addressed:**

† This paper is an extended version of our paper published in: R. Kosuru.; P. Chen.; S. Liu. A Reinforcement Learning based Power System Stabilizer for a Grid Connected Wind Energy Conversion System. In Proceedings of the 2020 IEEE Electric Power and Energy Conference (EPEC); IEEE: Edmonton, AB, Canada, November 9, 2020; pp. 1-5.

**Abstract:** Low frequency oscillations are of the primary issues for integrating a renewable source to the grid. The objective of this paper is to find the sensitive parameters that are causing low frequency oscillations and design a Twin Delayed Deep Deterministic Policy Gradient (TD3) agent controller to damp the oscillations without requiring the accurate system model. In this work, a Q-learning (QL) based model free wind speed DFIG is designed on the rotor-side converter (RSC) and a QL based model-free DC-link voltage regulator is designed on the grid-side converter (GSC) to enhance the stability of the system. In the next step, TD3 agent is trained to learn the system dynamics by replacing the inner current controllers from RSC which replaces the QL based model. In the first stage, the conventional PSS and Proportional Integral (PI) controllers are introduced to both RSC and GSC. Then, the system is trained to become model-free by replacing the PSS and the PI controller with QL algorithm under very small wind speed variations. In the second stage, the QL algorithm is replaced with TD3 agent by introducing large variable wind speeds. Results reveal that the TD3 agent can sustain the stability of the DFIG system under large variable wind speeds, without assuming a detailed control structure in prior, while QL based controllers can stabilize the doubly fed induction generator (DFIG) equipped wind energy conversion system (WECS) under small variable wind speeds.

**Keywords:** Q-Learning algorithm; Rotor Side Converter (RSC); Grid Side Converter (GSC); Power System Stabilizer (PSS); Small Signal (SS); Variable Speed Wind Energy Systems (VSWES).

1. Introduction

Based on the type of generator and the grid interface, wind energy systems can be categorized into four types [1]:

1. Squirrel cage induction generator (SCIG) or fixed speed system
2. Wound rotor induction generator (WRIG) with variable rotor resistance
3. Doubly fed induction generator
4. Full power converter generator

SCIG are mainly used for smaller wind turbines as they are simple and economical compared to other generators. In a squirrel cage induction generator, the rotor bars are permanently short circuited, therefore the rotor voltage is zero. The stator is connected to a soft starter and is connected to the grid through a transformer. A capacitor bank is employed to compensate for the reactive power and a soft starter is employed to mitigate high starting currents and to produce a smooth grid connection [2]. In WRIG, the stator is directly connected to the grid, and the wounded rotor winding is connected to a variable resistor via slip rings. In WRIG it is possible for the rotor to have configurations such as slip power recovery, the use of cyclo-converters, and rotor resistance chopper control [3]. In both WRIG and SCIG, by controlling the rotor resistance, the slip of the machine can be changed to 2-10% [1], through which the generator output can be controlled. The third type of generator used is the DFIG, constructional features of DFIG are like those of WRIG, except that in DFIG, the rotor winding of WRIG is connected to the grid through an AC-DC-AC converter. In fully converted synchronous generator, or fully converted squirrel cage induction generator (SCIG) the total power is interchanged between the wind system and the grid through a power electronic converter system whereas in some of the systems it can be transmitted to the grid directly [1]. Figure 1.1(a) and Figure 1.1(b) shows the configuration of a wind system with a synchronous generator and with a permanent magnet synchronous generator (PMSG). It can be recognized that in Figure 1.1(a), the synchronous generator is excited by using the power electronic converter externally, while in Figure 1.1(b), it is excited by a permanent magnet, as the generator is a permanent magnet synchronous generator. For large wind farms both DFIG and PMSG are preferred due to increased power control.



Figure1.1(a): Wind turbine system with full power converter generator (Synchronous Generator) [1]



Figure 1.1(b): Wind turbine system with full power converter generator (Permanent Magnet Synchronous Generator) [1]

There are many advantages and superior characteristics using a PMSG machine over DFIG. A PMSG machine has better performance, higher reliability, and wider speed control [4]. Due to PMSG’s better performance most of the current research work focus has been shifted to the topologies utilizing synchronous generators with permanent-magnet excited. However, DFIGs are still dominant in modern wind power generation systems, as DFIGs can operate under variable speed, regulation of active and reactive power independently, and low converter cost [5]. The reason for lower converter cost in DFIG wind turbines is due to the power electronic devices are fed with the power generated only from the rotor (25%-30% of rated power) and due to this the lower converter rating compromise offers a significant cost saving over PMSG [6].

The Doubly Fed Induction Generator (DFIG) model is considered as one of the best solutions for Wind Energy Conversion Systems (WECS). Two main reasons for using DFIG in WECS are its asynchronous characteristics and the flexibility of utilizing power electronic converters which results in cost saving due to lower converter rating. In DFIG based grid connected wind systems, the stator is directly connected to the grid whereas the rotor is connected to the grid by using a back-to-back power electronic converter. Figure 1.2 depicts the schematic model of a grid connected DFIG based WECS.

Every power network (grid) experiences a lot of disturbances, as there may be a great deal of variations like voltage, frequency, active power, and reactive power at the load end or at the power generating stations (renewable or non-renewable). For any system to operate without disturbances, power system stability and voltage regulation are the critical control issues that need to be considered [7]. The importance of power system stability can be illustrated clearly by investigating [8], which presents a clear description of power system instability as the primary grounds for any major black out. The main objective of any power system stabilizer is to provide stable power to the grid and to improve the damping of the oscillations. Especially when connecting any renewable energy conversion system to the grid, supplying a stable power is always challenging due to fluctuating frequencies and voltages.



Figure 1.2 Physical model of DFIG based wind generation system [9]

2. Literature Review

In a DFIG equipped WECS, the decoupled controllers are used to control the active and reactive power on both the rotor and grid side, the rotor speed, DC voltage and to track the maximum power. Low frequency oscillations have been observed for WECS with weak grid interconnections, this oscillation mode caused torsional interconnections with a remote synchronous generator and led to the shut down of the power plant [10]. To overcome the problem of low frequency oscillations a power system stabilizer can be applied at the output or the input of the controller. As per [11], a PSS can be employed for any DFIG variable that is influenced by the network oscillations such as rotor speed, stator electrical power, and voltage or network frequency. From [11]-[16], [17], [18], the importance of employing a PSS and improving the damping of the oscillations in a grid network fed by a DFIG-WES using slip signal and rotor speed deviations are discussed. Various oscillations that are damped using PSS are: electromechanical [19], inter-area power system oscillations [20], [21], network damping capability [22], oscillations caused by DFIG when integrated in a network that is already fed by a synchronous generator [17], [18]. From these papers it is clearly illustrated the need for a PSS to damp the oscillations and how a PSS is being used to improve the stability of the grid integrated wind system.

Most of the conventional PSS's (CPSSs) employed to wind energy systems are classical designs where the system is linearized around an operating point. However, there are PSS's developed using Artificial Intelligence (AI) techniques which are used for design and stability studies of the non-renewable sourced power systems. Some of the frequently used techniques for a PSS design are ANN [23, 24], and Fuzzy Logic [25, 10, 26] and support vector regression is used for designing an adaptive PSS in [27]. The use of AI to solve stability issues has been defined into 3 separate methodologies based on the techniques used: Supervised learning, Unsupervised learning, and Reinforcement learning (RL) [28]. RL is completely different from supervised learning and unsupervised learning. The supervised learning alone is not adequate for learning from interactions, and in interactive problems it is often impractical for supervised learning to obtain desired behaviour of all the situations in which the agent must act [28]. Whereas in an RL territory, an agent must be able to learn from its own experience. An unsupervised learning is typically about finding structure hidden in collections of unlabeled data, whereas the RL paradigm is about trying to maximize a reward signal instead of trying to find hidden structure [28]. So, an RL is a third machine learning paradigm, alongside supervised and unsupervised learning.

The current research has used reinforcement learning in controlling the entire system, as there is a need to interact between the mechanical system (wind turbine) and electrical system (DFIG and controllers) to supply a stable power to the grid for wind speed variations. One of the main reasons for using RL control method is, its capability of adopting itself to evolving generation levels, load levels, operating uncertainties, and repose to arbitrary disturbances [29]. In [29], RL method is used in online mode and applied to control a thyristor controller series capacitor (TCSC) aimed to damp power systems oscillations. RL controllers are designed to stabilize the closed loop system after severe disturbances in [30]. A specific RL algorithm called Q-Learning is utilized to control and adjust the gain of conventional PSS in [31]. RL algorithm is also used in generation control and voltage and reactive power control [32-34]. In [35] a control strategy is developed for PSS using Q-Learning method to suppress low frequency oscillations. In [36] a proportional resonance PSS (PR-PSS) is proposed using actor-critic agent, one of the RL techniques for adaptive adjustment of parameters to suppress ultra low frequency oscillations. The RL techniques discussed in [35] and [36] are used for comparing the PSS results obtained with the Q-Learning algorithm discussed in this chapter. TD3 method is implemeted in [37] and [38] for continuos power distribaces and to over come the low frequency oscillations. In [38], TD3 method is used for parameter estimation and fine tuning of PID controllers and overcome the problem of low frequency osciallations cased by load generation variations. Deep reinforcement learning methods are implemented for load frequency control of a multi-area power system and battery energy managmenet. In [39], a multi-agent deep reinforcement leanring method is proposed which utilized DDPG method for optimizing load ferequency control performance. In this paper in addition to Q-L method, TD3 method is explored and implemented to solve the low frequency oscillations caused by huge varying wind speeds. In this research, the TD3 method and Q-L method are implemented by replacing the existing PI controller and PSS. This paper is an extension to [40], where a Q-Leanring algorithm is implemented on rotor side converter for a slight range of change in wind speeds. The major contributions of the current work are discussed in the section 3.

3. Contributions

In this work, a CPSS firstly is designed, and the control strategies are applied at a fixed wind speed. The optimal PI gain values for both the rotor side converter and the grid-side converter are obtained by using eigenvectors for the PSS and the stability of the system is proved mathematically. In the next step, a Q-Learning algorithm is implemented over the designed PSS on RSC and PI controllers on the GSC for variable wind speeds. The objective of the Q-Learning algorithm implemented on RSC and GSC is to suppress the low frequency oscillations of the DFIG based WECS when the variable speeds are applied to the system. Since the control objective is to stabilize the system with active power (P) and reactive power(Q) without low frequency oscillations, this paper uses the active power change as the state of agent, and the control output of RSC as the action of the agent to train the model. On the grid side converter, the grid side active power is controlled which is a function of DC-link voltage () which acts as the state of agent, and the reward function generates the input signal to the grid side current (). Simulation results verify the designed Q-learning based model-free controllers can quickly stabilize the DFIG wind system under a small range of wind speeds. The terms action, state and agent used in this section are explained in detail in section 3 of this chapter. With Q-Learning agent, the system becomes unstable under huge varying wind speeds. One of the solutions is to implement the system with an actor-critic method. In this research, TD3 agent is trained to learn the system dynamics under huge varying wind speeds and control the active and reactive. Real time wind speed variations from Ottawa region are used as system input. The PI controller in the inner current control loop of RSC is replaced with the TD3 agent. From the results discussed in section 5, it is observed that the TD3 agent can mitigate the frequency changes under huge varying wind speeds and provide a stable power to the grid. In both QL and TD3 implementations the PSS and the PI controllers are removed, and the agent is trained to learn the system dynamics and suppress the low frequency oscillations.

4. Wind Turbine System Structure and Model

Figure 1.2 shows the physical model of the grid connected DFIG based wind turbine system. This benchmark power system model is first introduced in [9] to study the effect of shaft systems and low frequency oscillations by comparing the switching-level (SL) and Fundamental Frequency (FF) models. As illustrated in Figure 1.2 the system consists of mechanical and electrical models. Mechanical model has 3 different components (1) wind turbine, (2) gear box and (3) pitch controller, all 3 of these components together provide a complete drive train model which provides required mechanical energy for the generator. The electrical model consists of a DFIG generator of which the stator is directly connected to the grid and the rotor is connected through back-to-back IGBT based pulse width modulation converters.

***Design of Drive Train Model:***

The mathematical representation of drivetrain is formed by the turbine rotating mass, low speed shaft, the gear box, high speed shaft and generator rotating mass. The modeling of drivetrain is developed by neglecting the mechanical twisting and stresses as these are more related to mechanical design and studies. Moreover, for power system stability studies it is suggested in [41], to consider the turbine, gearbox and the generator as rigid disks and shafts as mass less torsional springs. A two-mass model is recommended for power system stability studies, this is because there is more possibility for the representation of shaft stiffness and inertia constants [42]. Figure 4.1 shows the physical representation of two mass model. The dynamics of the two-mass model can be obtained by applying Newton’s equation of motion for each mass. The equations obtained are [9,43]:

Where and is the shaft torque and it is given as:



Figure 4.1: DFIG shaft system, represented by a two-mass model [9]

In Figure 4.1 and are the turbine and generator rotor speed, respectively. and are the mechanical torque applied to the turbine and electrical torque respectively; and are the turbine inertia constant and generator inertia constant respectively; and are the damping coefficients of the turbine and generator respectively; is the internal torque of the model; is the damping coefficient of the shaft between the two masses; is the spring constant or shaft stiffness. Finally, is the gear ratio of the gear box.

***DFIG Model:***

For power system stability studies, the DFIG machine is modeled by neglecting stator transients as they do not affect the electromechanical oscillations [44] and by neglecting stator transients it is easier to solve for stator and grid equations [45]. Similarly, the rotor electrical transients are also neglected as the rotor winding is controlled by the fast-acting converters [46]. In addition, the other assumptions that are made for modeling the DFIG are: the skin effect, the saturation effect, and the iron losses (hysteresis and eddy currents) are neglected as these phenomena contribute more for conducting loss performance, transient fault analysis. For designing DFIG, synchronous reference frame is used. The reason for representing the DFIG model in synchronous rotating reference frame is because the axis model is convenient to conduct steady-state analysis and derive a small signal model [47]. The equivalent circuit of a DFIG in synchronously rotating reference frame is given in Figure 4.2.



Figure:4.2 Equivalent circuit of DFIG in synchronous reference frame



Figure:4.3 two axis reference frames for DFIG

In Figure 4.3 and axis is orthogonal to each other rotating at an angular velocity of. Whereas, and represent stator variables displaced by. For analyzing an induction machine variable associated with the rotor also needs to be transformed to axis reference frame. From Figure 4.3, and represent rotor variables displaced by rotating at an angular velocity of.

Stator model represented in axis reference frame.

Rotor model represented in axis reference frame.

Mathematical model including saturation effect is provided below [48].

Voltage equations of DFIG with flux saturation are:

Where

***Control Strategies:***

To develop the control strategies for the converters any one of the axes in two reference frames, rotating with synchronous speed is aligned with either the flux or the voltages of the stator. The two commonly used control strategies are the stator flux-oriented control and the stator voltage-oriented control. In the stator flux-oriented control, the -axis is aligned with the stator flux linkage vector [9]. This results in and. Whereas in stator voltage-oriented control, the-axis is aligned with the stator voltage linkage vector, resulting in and. For rotor side converter control the stator flux-oriented control is used and for grid side converter controller stator voltage-oriented control is adopted [9]. For this work the stator voltage-oriented control is implemented for both the rotor side converter controller and grid side converter controller as discussed in [49]. Compared to flux-oriented control, voltage-oriented control has the advantage of deriving the model and aligning with the stator voltage space vector using the measured phase voltages. Another advantage is that, typically the grid side converter is controlled using the stator voltage orientation, so by choosing the voltage-oriented control for rotor side controller the model would be simpler to implement. Usage of voltage-oriented control on both RSC and GSC can be observed in [49]. The objective of these control strategies is to decouple the control of active and reactive power [49].

***Rotor Side Controllers:***

The main objective of the rotor side converter controller is to control both the active and reactive power of the stator. It consists of two control loops; the inner control loop regulates the and axis rotor currents, whereas the outer loop controls the active and reactive power of the generator stator. By neglecting the transients in the stator flux linkages in stator voltage orientation and under an assumption that resistive drops are negligible it is observed that active power is independent from - axis rotor current. Under the same assumptions, for the rotor model the stator reactive power is totally dependent on the - axis rotor current and it is independent from - axis rotor current. It can be concluded that, in RSC the stator active and reactive power can be controlled independently by rotor - axis and - axis currents respectively which generates stator active and reactive power reference current outputs and respectively.

The generated reference currents are fed as inputs to the inner current loops of the rotor current controller. The final outputs of the rotor side converter controller are and . The controller is designed based on the rotor voltage models which can be expanded from equation and and expressed as [49]:

While designing the rotor side controller with decoupling elements; if the transients of the machine are neglected the derivative terms become zero. With this condition the - axis rotor voltage will be in terms of ; similarly, the - axis rotor voltage will be in terms of . With these elements a transfer function can be developed for the inner current control loops and from this the PI gains are obtained. From the mathematical models developed it is observed that the design is the same for both the control loops so the control design is also the same for both the inner current control loops. In these control loops and are the outputs of the rotor current controller loops and they are fed to the pulse width modulator of the converter along with the decoupled elements. Finally, these obtained signals from the modulator are fed to the converter circuit and this is connected to the DC link of the model.



Figure 4.4: Rotor side converter controller block

In Figure 4.4; and are the rotor currents and they are transformed from rotor three phase currents to by applying a transforming angle of where is the angle obtained from at grid frequency and is the rotor angle. is the stator active power and is the stator reactive power and these are obtained from and. Reactive power reference is given as, whereas stator active power reference is generated from maximum power tracking method, for this work a simple look up table is used which tries to obtain the reference stator power by plotting for power and speed. The other reference at control is the generator reference value which is set by the speed.



Figure 4.5: PLL model

***Grid-Side Controllers:***

To design the grid side converter controller first the grid model in the reference frame is transformed into two reference frames, by using a phase locked loop (PLL) which provides required transformation angle and the frequency for synchronizing the model with the grid. The design and operation of PLL is derived from [50] and it is shown in Figure 2.5. The grid model in reference frame is transformed to reference frame based on the transformation matrix from equation [50].

The transformation angle is the grid side converter terminal voltage angle and it is expressed as; where is given as where is the grid side converter terminal frequency. The measured phase angle by PLL is given as and the measured frequency is given as and this provides the transformation angle and the frequency to the grid side converter controller design. The obtained angle and frequency information is used for both and transformation at the grid end.

Initially for operating PLL the - axis voltage of stator is obtained by transformation technique. Once the grid side converter terminal voltage is transformed and the transformation angle is then the obtained equation is; where is the measured phase angle and it is expressed as.The grid side voltage frequency is obtained by the addition of the error signal processed in the PI controller and ; and it is given as:

By using this three phase PLL the grid equations can be transformed in to two reference frames. The transformed equations of the grid model are as follows:

In and and are the stator voltages; and are the grid filter resistance and inductance respectively; and are the total currents supplied to the grid in reference frame. Finally, and are the transformed voltages of the grid terminal. The model for the total current supplied to the grid i.e., and are given as the sum of the stator currents and grid side converter currents. The grid side converter voltages in reference frame can be expressed as:

Finally, the grid voltages are expressed as:

From the above derived grid side model, the grid side converter controller can be designed. The main objective of the grid side converter controller is to regulate the DC-link voltage and exchange the power between the rotor side converter and the grid. The other objective of this controller is to control the reactive power that is delivered to the grid, at the grid side converter. Like the rotor side converter controller, the grid side converter controller also consists of two cascaded control loops. The DC-link voltage and the reactive power are controlled by the outer control loop whereas the inner current control loop regulates the current components in the grid side converter. From the above discussed grid model, the active power and the reactive power at the grid can be expressed as:

On grid side active (In these equation by applying the synchronously rotating reference frame and aligning the - axis on the grid voltage vector the obtained results are and. Applying this to and yields:

From grid side active and reactive powers the outer control loops can be designed, the obtained result would be a function of DC-link voltage and the grid side converter current [9, 49]. By using this, an independent control loop is developed for DC-link voltage with as its output. Similarly, the grid side reactive power also can be controlled individually. From this outer loop is obtained. Similarly, for the inner control loop design the same reference frame is applied. The inner control loops are mainly designed as current control loops which use grid side currents as inputs. GSC control block is given in Figure 4.6.



Figure 4.6: Grid side converter controller [9]

***Grid side inner current control loop:***

With and the inner current control loop can be designed. In and are fed as the inputs at the outer end of the controller design and in is fed at the outer end. and act as decoupling elements for grid side converter controller. The final obtained equations are:

From and the PI controller can be designed, and it can be observed that both the controllers have similar structure. Due to very low sampling period and as the system requires fast response rate the controller is limited to a PI controller. Moreover, with the usage of power electronic devices and with huge varying wind speeds which generates lot of noise in the system adding a derivative controller would result in an undesirable simulation result.

5. Q Learning (QL) and Twin Delayed Deep Deterministic Policy Gradient (TD3)

Reinforcement Learning (RL) algorithms are focused on goal-directed learning from iterations which are mainly used to solve closed-loop problems. RL uses the actions from learning systems that influences the later inputs [28]. An RL algorithm consists of a discrete set of environment states , a discrete set of agent actions , and a set of scalar reinforcement signals . Here the agent interacts with the environment through action and the agent receives the current state as input and then the agent chooses an action to generate an output. The final goal of any RL algorithm is to increase the long-run sum of values of the reinforcement signals which can learn over the time by trial-and-error method and solve the problem [51].

The environment in reinforcement learning is fully observable and can be described into Markov decision process (MDP) and most of the RL problems are formalized as MDPs. In MDP, the action taken in the current state also affects the next state and not just the current state itself, so action plays a dominant role. Due to the action in the current state a return reward would be assigned for the corresponding state-action pair [28].

Of the various available RL algorithms, Q- learning algorithm is considered as simple and easier to implement due to its simple way for agents to learn and act optimally in controlled Markovian domains. The other main advantage of Q-learning algorithm is, it is exploration insensitive: Q values will converge to the optimal values, independent of how the agent behaves while the data is being collected [51]-[52]. With these advantages, this chapter is using a Q-learning algorithm to train the agent to suppress the oscillations and provide a stable power to the grid under variable wind speeds. Assuming the best action is taken initially, the Q-learning optimal value function is taken as [51].

In the above equation, is the expected discounted reinforcement of choosing action in state . Once the action is taken the agent would be given a reward , as per the effectiveness of the action by observing the resulting state of the environment. Here is the probability of action applied to state which change the state to . For each action executed the values will converge with probability 1 to and when the values are nearly converged to their optimal values, the agent will act greedily by taking the action with the highest value, from this greedy policy the optimal action is determined. At any time, step there is at least one action whose estimated value is greatest or optimal, choosing the greatest value is called greedy action [28]. is the discount factor which discounts the rewards exponentially in the future [51]. Typically, an agent will look up the Q-memory look up table which has state and action and is updated as per [34], [51-52]. The parameter in equation updates the Q-memory and affects the number of iterations [34].

Even though Q-learning method is simple to implement, the agent cannot learn under huge varying conditions, which is one of the observations from the current research work and a limitation of QL-agent. The other problem with Q-L method is that the Q-table must be limited to certain states and actions and the Q-table cannot be updated for large state-action space. So, the current research work has explored and implemented TD3 method both to overcome the Q-table limitation and the agent to adopt to large varying conditions.

TD3 method is one of the model-free policy-based deep reinforcement learning algorithms built on the DDPG method [53]. The objective of TD3 method is to increase the stability and performance by considering the function approximation error [53]. In an actor-critic setting the learning target is:

Where is the learning target, is receiving reward for every action, is the new state of the environment, is the discount factor, is the optimal policy and is the function approximator with parameters .

The stability and performance are increased by applying three modifications [53]:

1. Clipped Double Q-learning:

In this update the value target cannot introduce any additional overestimation over using the standard Q-learning target. With pair of actors and critics , the target update is

In this update the computational costs are reduced as a single actor is optimized with respect to , and the same target is used for updating .

1. Target networks and delayed policy updates:

In this update, the target network is used to reduce the error over multiple updates. The policy network is updated at a low frequency than the value network, to minimize the error. The modification is to update the policy and target networks after a fixed number of updates to critic [53]. So very few policy updates are made with this modification and the policy updates are not repeated to an unchanged critic.

1. Target policy smoothing regularization:

In this approach, the relationship between similar actions is forced explicitly by

modifying the training procedure which is done by fitting the value of a small area

around the target action. This would have the benefit of smoothing the value esti-

mate by bootstrapping off similar state-action value estimate [53]. The expectation

over actions is approximated by adding a small amount of random noise to the tar-

get policy where the noise is kept close to the original action.

The modified target update is:

6. Deep Reinforcement Learning based WECS

***Design of PSS:***

For this paper, the PSS is designed based on the transformation technique. The block diagram for PSS with transformation technique is shown in the Figure 6.1.



Figure 6.1: Block diagram for PSS with transformation technique

The signal is obtained from PLL frequency ():

The transformation is defined as:

The developed PSS with transformation technique is implemented on the inner current controller of RSC on both and axis control loops. The block diagram with transformation technique is given in Figure 6.2.



Figure 6.2: RSC with transformation PSS

***Q-Learning algorithm on RSC:***

The main motive of the Q-learning algorithm used is to suppress low frequency oscillations generated by variable wind speeds by searching for the best action in each state. In order to ensure continuous exploration and avoid tending to local optimum, a pursuit algorithm based on the learning automata algorithm is utilized to the action policy. Initially actions are determined with the uniform probability distribution. When the Q-table is updated, the probabilities of actions are updated as follows [54].

where denotes the probability with action-state pair is selected in iteration k, and represent the action exploration rate. After the effect of this algorithm, tends to for a sufficiently large and an optimal policy is obtained. The specific Q-learning based adaptive parameter algorithm proposed in this paper is summarized as follows:

|  |
| --- |
| Algorithm 1 Q-learning Based Adaptive Parameter in Rotor Side Algorithm |
| **For** each episode **do**  Initialize  Initialize  **For** each step of episode **do**  Choose *a* from *s* based on the current distribution  Take action *a*, observe *r*,  Update according to (6.1)  Update according to (6.4)    **End for**  **End for** |

Episode in algorithm 1 is defined as a notion of final time step, when the agent-environment interaction breaks naturally into subsequences [28].

Damping of frequency oscillations discussed in this chapter can be formulated as an MDP problem due to the future state of the closed loop controller is always dependent on the current state. Since the problems considered by reinforcement learning can be modeled as MDP models, thus we transform the adaptive parameter problem into a MDP model by designing a five-tuples as follows:

Design of state space : is a set of states which represent configurations of the system. It is assumed that all possible states are finite. In order to damp the low frequency oscillation of the power system and obtain stable power output, we use active power as the state information.

The states of the MDP are described as follows: state-space (in per-unit value) is discretized into the 11 spaces: , , , , , , , , , , . We can see that the state distribution is unbalanced around zero, because the deviation of the active power at the rated value is also unbalanced with wind fluctuation. The QL agent is trained with 10 states and 10 action pair. The state at iteration zero is chosen random and at each incremental step the state would be in one of the intervals. So, the states are divided into various sections based on the simulation results using PSS with PI controller where the controlled variable is monitored. In this case it is the active power which is a feedback parameter dependent on various other system parameters. The state space is divided into ten interval sates between an interval of -∞ to ∞. Each state consists of interval as the system is dynamic and continuous the states are chosen between the intervals. The observation from the small signal model and the simulation model with PSS is, the change is power is within the range of -0.5 to 0.01. The intervals for each state are chosen randomly however, the number of states is chosen based on the action.

Design of action space : is a set of actions which are executed by the agent to influence the environment. As mentioned before, the output of the agent should be the controller parameter at the rotor side. In this paper, we get the discrete action as follows: A= [-0.025 -0.02 -0.015 -0.01 -0.005 0.005 0.01 0.015 0.02 0.025]. The action space is chosen based on the output signals from the PSS controller which are replaced by the QL agent. These are the list of actions which are provided to the agent to determine which action to be chosen at its corresponding state. At state 1, the action chosen can be any of the action and not necessarily is action 1. The Q table, which is updated as per the algorithm 1, would choose the either the action with max Q or a random Q which is updated as per (6.1). Even though any action can be chosen, the trained agent would use the Q values from the trained model. So under varying wind speeds, the agent would know which action to be chosen based on the chosen state from the Q updated values.

is the state transition function that shows the distribution of the next state after executing an action ak to the environment with current state . Since the parameter variation model is unknown, we can use temporal-difference (TD) method to train the adaptive parameter policy. TD learning is a combination of Monte Carlo and dynamic programming ideas. Q-learning is an off-policy TD control algorithm [28]. In this paper, we apply a Q-learning algorithm to optimize the parameter of the rotor side controller.

Design of reward function : is the function that maps the state-action pair to a scalar which represents the immediate reward after applying an action to the environment with state . In this paper, . This means that the more the active power deviates from the reference value, the smaller the immediate reward, which prompts adjustment of the controller parameters so that the active power reaches the reference value. The parameter β used in (4.4) is used for updating the probability distribution. The value of β is 0.1. The value of α is 0.2.

The control application of QL on RSC is employed as per the Figure 6.3.



Figure 6.3: QL controller design on RSC

From the Figure 6.3 it can be observed that the PI controller which is developed for the PSS with transformation technique is completely replaced with Q-learning algorithm.

***Q-Learning algorithm for DC-Link Voltage Control on GSC:***

As discussed in the above section, the primary objective of the Q-learning algorithm used is to suppress low frequency oscillations generated by variable wind speeds by searching for the best action in each state. On the GSC, the sensitive parameters are the DC-link voltage and grid side current which are observed from the small signal model. For the action policy, a pursuit algorithm based on the learning automata algorithm is utilized. Initially actions are determined with the uniform probability distribution. When the Q-table is updated, the probabilities of actions are updated using (6.4). The Q-learning algorithm discussed for RSC (Algorithm 1) is used for updating the episodes on GSC as well, the main difference is with the action and reward that is chosen for each controller. For GSC the action space is designed for control parameters and and the reward function is defined from DC-link voltage. The new controller design with QL on grid side converter is shown in Figure 6.4.



Figure 6.4: QL controller design on GSC

It can be observed that the PI controllers on GSC are replaced with the QL algorithm. The adaptive parameter problem can be transformed into a MDP model by designing a five-tuples as follows:

Design of state space : To damp the low frequency oscillation the sensitive parameters observed from the GSC are the DC-link voltage and input reference current to the current control loop, so DC-Link voltage is used as the state information.

The states of the MDP are described as follows: On the GSC, the same state spaces defined for RSC can be used as the objective is to provide a stable power to the grid which is a function of active power from RSC and GSC. And as well, the deviation of the active power at the grid side converter is a function of which is also unbalanced with wind fluctuation. However, the action space would be completely different from the RSC action spaces as the control parameters differ from each other.

Design of action space : The output of the agent is a controller parameter and at grid side. In this paper, we get the discrete action as follows:

A= [-0.19 -0.186 -0.18 -0.176 -0.172 -0.17 -0.168 -0.164 -0.16 -0.15]

The action space on GSC is chosen based on the PI controllers which are replaced with the QL-agent. However, the state space would remain the same as the end output is the active power which is monitored and controlled by the agent.

is the state transition function that shows the distribution of the next state after executing an action a k to the environment with current state.

Design of reward function : In this paper, . This means that the more the grid side active power deviates from the reference value, the smaller the immediate reward, which prompts adjustment of the controller parameters so that the voltage reaches the reference value.

A complete block diagram with QL-controller units discussed above can be observed from Figure 6.5.



Figure 6.5: Block diagram with QL controller

In Figure 6.5, is the wind speed, is the generator rotor speed, is the rotor angle which is used for transformation from reference frame for rotor currents. and are stator voltage and currents respectively, whereas and are grid voltage and currents, respectively. Grid voltage is provided as input to the phase locked loop, which extracts the angle .The is used for transformation from reference frame for stator voltage, stator current, grid voltage and grid currents. From Figure 6.5, it can be observed that voltages fed to the rotor side converter and grid side converter are provided by the QL- controllers which are discussed in this section.

***TD3 method:***

As discussed in section 5, TD3 method is derived from DDPG method with few modifications. In this section TD3 algorithm and its implementation is discussed. One of the objectives of this research work is to control the frequency under huge varying speeds by replacing the PI controllers, as the PI controllers cannot provide or produce satisfying tracking performance with huge varying wind speeds. In which, TD3 method is implemented which provides a nonlinear control. In this paper, the inner current PI controllers are completely replaced with TD3 algorithm. TD3 agent is implemented with an environment where the observation space is continuous or discrete and action space is continuous. In this experience the both the observation and action space are continuous. When it comes to actor and critic, the actor is deterministic policy actor, and the critic can be one or more Q-value function critics . The training process consists of there steps. In step 1, the actor and critic properties are updated at each iteration by the agent. In step 2, experience buffer is used to store the past experiences. From the experience buffer the actor and critic uses a mini batch of experiences randomly. In step 3, a noise is given to the action chosen by the policy using the stochastic noise model at each episode.

TD3 implementation has four stages. In stage 1, the actor and critic functions are chosen. The actor is a deterministic actor that returns the action which maximizes the long-term reward with input with parameters and state . Apart from the actor to choose the best action, a target actor is also developed to improve the stability, the target actor parameters are updated using the latest actor parameters. In critics there are two critics, one is the value critic, and the second critic is the target critic. The value critic takes the state and action as inputs and provides the expectation of the long-term reward. The target critic is responsible to improve the stability of the optimization. Target critic parameters are updated using the latest critic parameter values. In stage 2, agent is created with state and action specification of the environment. As the agent has both actor and critic networks with the environment, now in stage 3, the agent is trained to learn and update the actor and critic models at each episode. The training is implemented as per algorithm 2 [53]. In stage 4, the target actor and the target critic parameters are updated using one of the target update methods like periodic, smoothing, and periodic smoothing. However, for this research work smoothing update is chosen.

|  |
| --- |
| Algorithm 2 TD3 [53] |
| Initialize critic networks with random parameters  Initialize target critic with same random parameters ; so  Initialize actor network with random parameters  Initialize target actor network with same random parameters  So, for target networks  Initialize replay buffer  **For** t = 1 to T do  For current state of observation select action with exploration noise . Here is the stochastic noise from the noise model  Execute action and observe reward and new state .  Store the experience in (experience buffer)  Sample a random mini batch of transitions from  If is a terminal state, set the value function target else  Update critics  **If** mod then  Update by deterministic policy gradient:    Update target networks (smoothing):      **end if**  **end for** |

In the current work, the inner current controller on the rotor side controller is replaced with TD3 agent. So, the inputs for TD3 agent are . Here are rotor side currents and are the output of the outer current control loop on RSC. and are the rotor speed and reference speed respectively. All the current and speed with their references is provided as input or are used as the parameters for observation. The reward is computed by taking the error between the reference values and the actual values for currents. The reward is computed as shows in equation (6.5).

So, the number of observations is six which are the inputs to the TD3 agent, and the number of actions is two and . As we have the observations and action space available, the next step is to create the agent block, as shown in Figure 6.6.



Figure 6.6: TD3 agent implementation on RSC

For the TD3 agent, two critic networks are created with deep neural network with state and action as inputs and one output. The neural network developed is a fully connected layer for both state and action path. In the same way for both the actor networks fully connected layers are used. TD3 agent will determine the action need to be taken for the given state. To train the agent, the TD3 agent is provided with its discount factor, the buffer length, mini batch size, target smoothing factor and finally the target update frequency. For training, the agent is allowed to run each training for 1000 episodes and stop training when the agent receives a cumulative average reward >= -200 over 100 consecutive episodes. Training progress with TD3 agent is shown in Figure 6.7 which shows the average reward of each episode increases and policy becomes stable after 40episodes. The training process is executed for a duration of 32 hours, as it reached the specified maximum episode numbers.

Chart

Description automatically generated

Figure 6.7: TD3 agent training progress

7. Results and Discussion

***Simulation results with the newly designed PSS:***

The model discussed in this chapter is built in MATLAB/Simulink. The topology is shown in Figure 6.5. There have been extensive works published based on simulations [10], [4], [35], [55]. A small signal model is developed, to identify the state variables that affect the stability of the overall system. From a small signal model, the sensitive variables observed are,,, , and . The input given to the power system stabilizer is the frequency. One of the main observations from small signal stability analysis is that the state variables associated with the inner current control loop tend to move faster towards the real axis making the system unstable. This is observed with the large signal model as well by applying small faults. Performance evaluation of the system with PSS and without PSS can be observed from the small signal model. The low frequency mode of the system can be observed from the same small signal model. Usage of Eigen value distribution of systems with and without PSS to identify ultra-low frequency oscillation mode can be observed in [36]. The shift of the eigenvalues with the new PSS implemented is noted in the table below. The highlighted eigenvalues of are newly added signal input from the PSS. The time constants for the washout filter and compensator are determined from the small signal model by observing the stability of the system. and are the signal input of the outer current control loop of the grid side converter. and are the input signals to the PI controllers at the inner current control loop of RSC. is the input signal to the PI controller at the inner current control loop of GSC.

Table 7.1: Small signal model Eigen values and state variables

|  |  |  |
| --- | --- | --- |
|  |  | State Variables |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Results for power system stabilizer with transformation technique:



Figure 7.1: (a). Active power and (b). Reactive power with and without PSS

From the graph it can be observed that there is an increase in the damping of the oscillations for both active power and generator speed with a PSS implemented with transformation technique.



Figure 7.2: (a). DC-link voltage and (b). Generator speed with PSS

***Fault analysis with transformation PSS:***

The plot with transformation PSS during fault condition. It can be observed that the oscillations are damped even during the fault condition. This can be more clearly observed from the power plot and in which the oscillations are damped the settling time is also very less. From this discussion it can be concluded that the PSS is working effectively during the fault condition.



Figure 7.3: (a). Active power and (b). Reactive power with and without PSS during three phase fault condition

For fault analysis fault analysis, a three-phase fault is applied to the system for a time of between times to. The results obtained for active power and reactive power are shown on Figure 7.3 (a) and (b) respectively. From the above plot it can be observed that the PSS with voltage as input is able to damp the oscillations during the fault and the settling time after fault is also very less.

|  |  |
| --- | --- |
|  |  |
| Figure 7.4: (a). Active power and (b). Reactive power with and without PSS during single phase short circuit fault condition | |

Figure 7.4 provides the simulation results for system under single phase short circuit fault. The fault is applied to the system for between times to. It can be observed that the PSS is able to damp the oscillations.

***Results with Q-learning algorithm:***

Now a Q-learning algorithm is implemented on the developed PSS at RSC and outer current control loop of the GSC, with a variable wind speed. The algorithm model is applied at and . Since the control objective is to stabilize the output of the power system with active and reactive power without low frequency oscillations, this paper uses the parameters that are directly related to the active power change of the power system as the state of the agent, and the control output of the controller as the action of the agent to train the appropriate control Strategy. After testing, when the wind speed fluctuates in a small range, it is easier to stabilize the system through the reinforcement learning controller. Therefore, the wind speed of this experiment is designed to be between14 m/s to 15 m / s.

The following is the output waveform of active power of the power system after several iterations of the reinforcement learning controller. The active power waveform is observed at iteration 18, which can be observed from Figure 7.5, respectively.

A picture containing graphical user interface

Description automatically generated  
Figure 7.5: Active power response at 18th iteration

From Figure 7.5, it can be observed that the generated is tending to be smooth without any oscillations, which indicates that the QL algorithms implemented were able to learn from the system and provide the desired output.

Chart

Description automatically generated

Figure 7.6: Active power response at 25th iteration

However, the final active power is still not smooth. Next the system is iterated until the desired active power is generated without any oscillations, which is observed from 21st iteration. Figure 7.6 shows the response of the active power at 25th iteration. After this point the system response will not differ, as the Q-learning agent is able to learn about the system at various states and provide required action at different states. From Figures 5.6 to 5.8, it is observed that under the action of the trained reinforcement learning controller, the system can output stable active power without low frequency oscillation. The total time taken for 25 episodes or iterations is 90 minutes.

Chart

Description automatically generated

Figure 7.7: Reactive power with QL

Figure 7.7 shows the control of reactive power at 25th iteration. It can be observed that the QL agent is able to learn from the environment and control reactive power as well.

When the wind speed changes, the output speed and power graph is shown in Figure 7.8. As can be seen from Figure, 7.8(a) when the wind speed changes, the speed and active power of the power system are stabilized.

Chart, line chart

Description automatically generated

Figure 7.8: (a) Wind speed change (b) Generator speed in per unit

***Comparing Q-learning algorithm with PI controllers:***

The Q-learning algorithm developed is completely replacing the PI controllers both on RSC and GSC and the power system stabilizer at RSC. From the results the observation is, the model free algorithm can learn from the system environment and generate similar results as of the PI controller.

Chart

Description automatically generated

Figure 7.9: Active power with PI controller vs Q-learning

In Figure 7.9, the dotted red line indicates the cure for active power with QL algorithm and the continuous curve shows the active power with PI controllers. It can be observed that the Q-learning algorithm is controlling the active power like that of the PI controller. The advantage of the developed Q-learning model over PI controller design is, the Q-learning agent is providing the controller parameters dynamically whereas the gains in PI controller are fixed and moreover the Q-learning algorithm us implemented with varying wind speeds whereas the PI controller is implemented with a constant wind speed. Figure 7.10 shows a comparison for reactive power for the developed PSS with PI controller and Q-learning algorithm.

Chart

Description automatically generated

Figure 7.10: Reactive power with PI controller vs Q-learning

***Comparing Q-learning algorithm with PI controllers under fault conditions:***

To evaluate the control of Q-learning algorithm under fault conditions a three-phase fault is employed at the grid end before the transformer. Clearing time is 2 sec, starts at 30 sec and clears at 32 sec. Figure 7.11 and Figure 7.12 shows the active power and reactive power plot respectively under the fault conditions. In Figure 7.11, the red dotted line is the active power with Q-learning algorithm and the continuous line shows the curve for PI controller. It can be observed that the damping of oscillations is more effective with Q-learning algorithm when compared to the PSS with PI controllers.

Chart

Description automatically generated

Figure 7.11: Active power with PI controller vs Q-learning under fault

In Figure 7.12, the red dotted curve indicates the reactive power with Q-Learning algorithm, and the continuous curve is with PI controller. From the figure it can be observed that the damping of oscillation is more effective with Q-leanirng algorithm when compared with the PI controllers.

Chart

Description automatically generated

Figure 7.12: Reactive power with PI controller vs Q-learning under fault

The proposed Q-learning algorithm can be compared with the PSS using Q-learning algorithm from [35]. In [35], the case study used is a four machine two area system where a Q-learning based PSS is employed to damp the frequency oscillations. The same can be observed in the current chapter, however the systems used are completely different. Moreover, in the current chapter the performance of the Q-learning is observed under fault conditions as well. From Figure 7.9 to 7.12 a clear comparison can be observed between the PSS with PI controller and the Q-Learning. Here PSS with PI controller can be treated as a test case to evaluate the performance of the Q-learning based model. From Figures 7.9-7.12 the proposed Q-learning based model reaches steady state in the shortest time compared with the classical PI controller-based PSS.

***Comparing TD3 Agent with Q-learning algorithm:***

Simulations are carried out with large variable wind speed to PI, Q-learning and TD3 agent. The wind speeds used are from the Ottawa region for the complete month of June 2019.

Figure 7.13: Wind speed variation in Ottawa region

Figure 7.14: Average wind speed variation in Ottawa region

The wind speeds variation shown in Figure 7.13 is used as an input to the DFIG WECS. With the varying wind speeds, it is hard for the regular PI controller to maintain the frequency and at the same to provide a table output to the grid. The average wind speed variations can be observed from Figure 7.14. The active power, reactive power with PSS with QL vs PSS with TD3 agent can be seen from Figure 7.15 and Figure 7.16 respectively. It can be observed that the both active power and reactive power are not stable with such huge wind speed variations in both PSS with PI controller and Q-learning agent. However, once the TD3 agent implemented on the inner current control loop of the RSC learns the system dynamics, the agent can control the speed generator speed and inner currents with respect to their reference values. So, with large variable wind speeds and without using any inner PI controller, with TD3 agent the system can produce stable active and reactive power as per the varying wind speed by mitigating the lower frequencies. The system with PI controller doesn’t even respond to large varying wind speeds, as it can work only for a constant wind speed.

Chart, line chart

Description automatically generated

Figure 7.15: Active power with TD3 agent vs Q-learning

Chart, line chart

Description automatically generated

Figure 7.16: Reactive power with TD3 agent vs Q-learning

Graphical user interface, text, application

Description automatically generated Figure 7.17: DC-Link voltage with TD3 agent vs Q-learning

Chart, line chart, histogram

Description automatically generated

Figure 7.18: Generator speed with TD3 agent vs Q-learning

Figure 7.17 and Figure 7.18, shows the DC-link voltage and generator speed comparison between QL agent and TD3 agent. From the figures, it can be observed that TD3 agent is can learn and control the system under large varying wind speed and performs better than the QL agent.

8. Limitation and Future work

The wind DFIG system with PSS and PI controllers works fine for a constant speed. Which is observed under nor mal and fault conditions. If a varying wind speed is provided, then the PSS cannot suppress any oscillations and system is very unstable. PSS with PI controller cannot be adopted or used for large varying wind speeds. To overcome the limitation of varying wind speeds, reinforcement learning is used to learn the system dynamics and mitigate the oscillations. For this first the PSS and the PI controllers both from RSC and GSC are replaced with QL-agent. The Q-learning algorithm developed in this paper is implemented for a very short range of varying wind speeds and does not respond accurately for large varying wind speeds. The observation is that the QL agent is not successful in learning the system dynamics with huge varying wind speeds. To overcome the limitation of application of an RL to large varying wind speeds, advanced RL techniques like actor-critic (A2C or A3C) or policy gradient methods like Deep Deterministic Policy Gradient (DDPG) will be helpful. As these methods are robust in learning about the environment. To handle large varying wind speeds, TD3 method is introduced in this research. Usage of A3C based strategy can be observed in [36], where the PSS parameters are tuned to suppress the low frequency oscillations for a 10-machine 39-bus transmission network. So, an A2C or A3C can be adopted to suppress low frequency oscillations for a large varying wind speeds in DFIG based WECS. As discussed at the beginning of section 7, the complete model is built in the MATLAB/Simulink environment. For this research work, there is no experimental hardware involved. The complete results with comparison are discussed in section 7.

9. Conclusions

In this chapter a DFIG equipped WECS is developed, and closed loop controllers are designed to damp the oscillations by controlling both active and reactive power. First a small signal model is developed to identify the sensitive parameters that affect the stability of the system and from the SS model the proportional and integral gains are derived. The derived PI gains are used in the inner current control loops of the large signal model. Next, the PI controllers are completely replaced with the Q-learning based RL technique and the Q-learning based model-free power system stabilizer and DC-link voltage regulator are developed on both RC and GSC. From the results it is observed that the designed model free PSS can damp the oscillations under small varying wind speeds and the faulty conditions. The conclusion is that the QL algorithm agent can learn from the environment and control the active power by helping the system to operate under normal conditions with variable wind speed, by making the system model free. However, the limitation with QL method is, it cannot control under large varying wind speeds. To overcome this limitation, an actor-critic method called TD3 method is introduced. From the results it can be observed that TD3 agent can learn the system dynamics under large varying wind speeds and agent can deliver desired active and reactive power to the grid.

**Conflicts of Interest:** Declare conflicts of interest or state “The authors declare no conflict of interest.” Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state “The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results”.

**Appendix A:** Parameters for the system:

***Base quantities:***

– Generator rated power

– Stator rated voltage

– Electrical base speed

–Base current

– Base impedance

–Base inductance

– Base flux

– Base active and reactive power

***DFIG Parameters:***

– Synchronous speed

- Stator resistance

–Rotor resistance

- Stator inductance

- Rotor inductance

- Mutual inductance

- Generator inertia constant

- No of pairs of poles

Drive train data:

- Wind turbine inertia constant

- Shaft spring constant

- Shaft mutual damping

- rated wind speed

Blade length

– Air density

Turbine rated speed

Tip speed ratio

- Maximum value of

; ; ; ; ; ; ; ;

DC-Link:

***Controller data:***

Rotor side converter controller:

– Active power loop

– Inner current controller loop

– Stator Reactive power loop

– Inner current controller loop

Grid side converter controller:

–DC-link controller

– Inner current controller loop

– Grid side Reactive power controller

– Inner current controller loop

PLL:

00

***Reference values:***

Reactive power reference values

PSS with voltage as input:

;

PSS with frequency as input (transformation technique):

;

; And.

Pitch controller:

; ; ;

**References:**

1. Ellis, A.; Muljadi, E. Wind Power Plant Representation in Large-Scale Power Flow Simulations in WECC. In Proceedings of the 2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century; IEEE: Pittsburgh, PA, USA, July 2008; pp. 1–6.
2. Zobaa, A.F.; Bansal, R.C. *Handbook of Renewable Energy Technology*; WORLD SCIENTIFIC, 2011; ISBN 978-981-4289-06-1.
3. Nayar, C.V.; Islam, S.M.; Dehbonei, H.; Tan, K.; Sharma, H. Chapter 1—Power Electronics for Renewable Energy Sources. In *Alternative Energy in Power Electronics*; Rashid, M.H., Ed.; Butterworth-Heinemann: Oxford, UK, 2011; pp. 1–79.
4. Zhou, F.; Liu, J. A Robust Control Strategy Research on PMSG-Based WECS Considering the Uncertainties. *IEEE Access* **2018**, *6*, 51951–51963, doi:[10.1109/ACCESS.2018.2870411](https://doi.org/10.1109/ACCESS.2018.2870411).
5. Shang, L.; Hu, J. Sliding-Mode-Based Direct Power Control of Grid-Connected Wind-Turbine-Driven Doubly Fed Induction Generators Under Unbalanced Grid Voltage Conditions. *IEEE Trans. Energy Convers.* **2012**, *27*, 362–373, doi:[10.1109/TEC.2011.2180389](https://doi.org/10.1109/TEC.2011.2180389).
6. Hui Zeng; Yu Zhu; Jinsong Liu Verification of DFIG and PMSG Wind Turbines’ LVRT Characteristics through Field Testing. In Proceedings of the 2012 IEEE International Conference on Power System Technology (POWERCON); IEEE: Auckland, October 2012; pp. 1–6.
7. Bevrani, H.; Watanabe, M.; Mitani, Y. *Power System Monitoring and Control*; Wiley: Hoboken, New Jersey, 2014; ISBN 978-1-118-85247-7.
8. Vassell, G.S. Northeast Blackout of 1965. *IEEE Power Eng. Rev.* **1991**, *11*, 4, doi:[10.1109/MPER.1991.88621](https://doi.org/10.1109/MPER.1991.88621).
9. Qiao, W. Dynamic Modeling and Control of Doubly Fed Induction Generators Driven by Wind Turbines. In Proceedings of the 2009 IEEE/PES Power Systems Conference and Exposition; IEEE: Seattle, WA, USA, March 2009; pp. 1–8.
10. Dehghani, M.; Han, W.; Karimipour, H. Coordinated Fuzzy Controller for Dynamic Stability Improvement in Multi-Machine Power System. In Proceedings of the 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE); IEEE: Oshawa, ON, August 2018; pp. 165–170.
11. Anaya-Lara, O. Power System Stabiliser for a Generic DFIG-Based Wind Turbine Controller. In Proceedings of the 8th IEE International Conference on AC and DC Power Transmission (ACDC 2006); IEE: London, UK, 2006; Vol. 2006, pp. 145–149.
12. Hughes, F.M.; Anaya-Lara, O.; Jenkins, N.; Strbac, G. Control of DFIG-Based Wind Generation for Power Network Support. *IEEE Trans. Power Syst.* **2005**, *20*, 1958–1966, doi:[10.1109/TPWRS.2005.857275](https://doi.org/10.1109/TPWRS.2005.857275).
13. Hughes, F.M.; Anaya-Lara, O.; Jenkins, N.; Strbac, G. A Power System Stabilizer for DFIG-Based Wind Generation. *IEEE Trans. Power Syst.* **2006**, *21*, 763–772, doi:[10.1109/TPWRS.2006.873037](https://doi.org/10.1109/TPWRS.2006.873037).
14. Hughes, F.M.; Anaya-Lara, O.; Ramtharan, G.; Jenkins, N.; Strbac, G. Influence of Tower Shadow and Wind Turbulence on the Performance of Power System Stabilizers for DFIG-Based Wind Farms. *IEEE Trans. Energy Convers.* **2008**, *23*, 519–528, doi:[10.1109/TEC.2008.918586](https://doi.org/10.1109/TEC.2008.918586).
15. Mishra, Y.; Mishra, S.; Fangxing Li; Zhao Yang Dong; Bansal, R.C. Small-Signal Stability Analysis of a DFIG-Based Wind Power System Under Different Modes of Operation. *IEEE Trans. Energy Convers.* **2009**, *24*, 972–982, doi:[10.1109/TEC.2009.2031498](https://doi.org/10.1109/TEC.2009.2031498).
16. Mishra, Y.; Mishra, S.; Tripathy, M.; Senroy, N.; Dong, Z.Y. Improving Stability of a DFIG-Based Wind Power System With Tuned Damping Controller. *IEEE Trans. Energy Convers.* **2009**, *24*, 650–660, doi:[10.1109/TEC.2009.2016034](https://doi.org/10.1109/TEC.2009.2016034).
17. Surinkaew, T.; Ngamroo, I. Coordinated Robust Control of DFIG Wind Turbine and PSS for Stabilization of Power Oscillations Considering System Uncertainties. *IEEE Trans. Sustain. Energy* **2014**, *5*, 823–833, doi:[10.1109/TSTE.2014.2308358](https://doi.org/10.1109/TSTE.2014.2308358).
18. Iswadi, H.R.; Morrow, D.J.; Best, R.J. Small Signal Stability Performance of Power System during High Penetration of Wind Generation. In Proceedings of the 2014 49th International Universities Power Engineering Conference (UPEC); IEEE: Cluj-Napoca, Romania, September 2014; pp. 1–6.
19. Mendonca, A.; Lopes, J.A.P. Simultaneous Tuning of Power System Stabilizers Installed in DFIG-Based Wind Generation. In Proceedings of the 2007 IEEE Lausanne Power Tech; IEEE: Lausanne, Switzerland, July 2007; pp. 219–224.
20. Elkington, K.; Ghandhari, M.; Söder, L. Using Power System Stabilisers in Doubly Fed Induction Generators. *Australasian Universities Power Engineering Conference* **2008**, 6.
21. Ke, D.P.; Chung, C.Y.; Xue, Y. Controller Design for DFIG-Based Wind Power Generation to Damp Interarea Oscillation. In Proceedings of the 2010 5th International Conference on Critical Infrastructure (CRIS); IEEE: Beijing, China, September 2010; pp. 1–6.
22. Gong, B.; Xu, D.; Wu, B. Network Damping Capability of DFIG-Based Wind Farm. In Proceedings of the 2010 IEEE Energy Conversion Congress and Exposition; IEEE: Atlanta, GA, September 2010; pp. 4083–4090.
23. Zhang, Y.; Chen, G.P.; Malik, O.P.; Hope, G.S. An Artificial Neural Network Based Adaptive Power System Stabilizer. *IEEE Trans. On energy Conversion* **1993**, *8*, 71–77, doi:[10.1109/60.207408](https://doi.org/10.1109/60.207408).
24. Kahouli, O.; Alshammari, B.; Dhouib, B. Application of ANN and ANFIS Techniques for PSS Tuning in a Multimachine Power System. In Proceedings of the 2019 16th International Multi-Conference on Systems, Signals & Devices (SSD); IEEE: Istanbul, Turkey, March 2019; pp. 435–440.
25. Hariri, A.; Malik, O.P. A Fuzzy Logic Based Power System Stabilizer with Learning Ability. *IEEE Trans. On energy Conversion* **1996**, *11*, 721–727, doi:[10.1109/60.556370](https://doi.org/10.1109/60.556370).
26. Chaturvedi, D.K.; Malik, O.P. Neurofuzzy Power System Stabilizer. *IEEE Trans. On Energy Conversion* **2008**, *23*, 887–894, doi:[10.1109/TEC.2008.918633](https://doi.org/10.1109/TEC.2008.918633).
27. Boonprasert, U.; Theera-Umpon, N.; Rakpenthai, C. Support Vector Regression Based Adaptive Power System Stabilizer. In Proceedings of the Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS ’03.; IEEE: Bangkok, Thailand, 2003; Vol. 3, p. III-371-III–374.
28. R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
29. Ernst, D.; Glavic, M.; Wehenkel, L. Power Systems Stability Control: Reinforcement Learning Framework. *IEEE Trans. Power Syst.* **2004**, *19*, 427–435, doi:[10.1109/TPWRS.2003.821457](https://doi.org/10.1109/TPWRS.2003.821457).
30. Hadidi, R.; Jeyasurya, B. Reinforcement Learning Based Real-Time Wide-Area Stabilizing Control Agents to Enhance Power System Stability. *IEEE Trans. Smart Grid* **2013**, *4*, 489–497, doi:[10.1109/TSG.2012.2235864](https://doi.org/10.1109/TSG.2012.2235864).
31. Hadidi, R.; Jeyasurya, B. Reinforcement Learning Approach for Controlling Power System Stabilizers. *Can. J. Electr. Comput. Eng.* **2009**, *34*, 99–103, doi:[10.1109/CJECE.2009.5443857](https://doi.org/10.1109/CJECE.2009.5443857).
32. Imthias Ahamed, T.P.; Nagendra Rao, P.S.; Sastry, P.S. A Reinforcement Learning Approach to Automatic Generation Control. *Electric Power Systems Research* **2002**, *63*, 9–26, doi:[10.1016/S0378-7796(02)00088-3](https://doi.org/10.1016/S0378-7796(02)00088-3).
33. Yu Tao; Zhou Bin A Novel Self-Tuning CPS Controller Based on Q-Learning Method. In Proceedings of the 2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century; IEEE: Pittsburgh, PA, USA, July 2008; pp. 1–6.
34. Vlachogiannis, J.G.; Hatziargyriou, N.D. Reinforcement Learning for Reactive Power Control. *IEEE Trans. Power Syst.* **2004**, *19*, 1317–1325, doi:[10.1109/TPWRS.2004.831259](https://doi.org/10.1109/TPWRS.2004.831259).
35. Zhu, X.; Jin, T. Research of Control Strategy of Power System Stabilizer Based on Reinforcement Learning. In Proceedings of the 2020 IEEE 2nd International Conference on Circuits and Systems (ICCS); IEEE: Chengdu, China, December 10 2020; pp. 81–85.
36. Zhang, G.; Hu, W.; Cao, D.; Huang, Q.; Yi, J.; Chen, Z.; Blaabjerg, F. Deep Reinforcement Learning-Based Approach for Proportional Resonance Power System Stabilizer to Prevent Ultra-Low-Frequency Oscillations. *IEEE Trans. Smart Grid* **2020**, *11*, 5260–5272, doi:[10.1109/TSG.2020.2997790](https://doi.org/10.1109/TSG.2020.2997790).
37. Li, J.; Yu, T. Deep Reinforcement Learning Based Multi-Objective Integrated Automatic Generation Control for Multiple Continuous Power Disturbances. *IEEE Access* **2020**, *8*, 156839–156850, doi:[10.1109/ACCESS.2020.3019535](https://doi.org/10.1109/ACCESS.2020.3019535).
38. Khalid, J.; Ramli, M.A.M.; Khan, M.S.; Hidayat, T. Efficient Load Frequency Control of Renewable Integrated Power System: A Twin Delayed DDPG-Based Deep Reinforcement Learning Approach. *IEEE Access* **2022**, *10*, 51561–51574, doi:[10.1109/ACCESS.2022.3174625](https://doi.org/10.1109/ACCESS.2022.3174625).
39. Yan, Z.; Xu, Y. A Multi-Agent Deep Reinforcement Learning Method for Cooperative Load Frequency Control of a Multi-Area Power System. *IEEE Trans. Power Syst.* **2020**, *35*, 4599–4608, doi:[10.1109/TPWRS.2020.2999890](https://doi.org/10.1109/TPWRS.2020.2999890).
40. Kosuru, R.; Chen, P.; Liu, S. A Reinforcement Learning Based Power System Stabilizer for a Grid Connected Wind Energy Conversion System. In Proceedings of the 2020 IEEE Electric Power and Energy Conference (EPEC); IEEE: Edmonton, AB, Canada, November 9 2020; pp. 1–5.
41. J.F.Manwell, J.G.Mcgowan and A.L.Rogers. *Wind Energy Explained, theory, design, and applications*. John Wiley & Sons Ltd: Chichester, U.K, 2009.
42. Salman, S.K.; Teo, A.L.J.; Rida, I.M. The Effect of Shaft Modelling on the Assessment of Fault CCT and the Power Quality of a Wind Farm. In Proceedings of the Ninth International Conference on Harmonics and Quality of Power. Proceedings (Cat. No.00EX441); IEEE: Orlando, FL, USA, 2000; Vol. 3, pp. 994–998.
43. Mei, F.; Pal, B.C. Modelling and Small-Signal Analysis of a Grid Connected Doubly-Fed Induction Generator. In Proceedings of the IEEE Power Engineering Society General Meeting, 2005; IEEE: San Francisco, CA, USA, 2005; pp. 1503–1510.
44. P. Kundur, *Power System Stability and Control*. McGraw-Hill: New York, USA, 1994.
45. Ledesma, P.; Usaola, J. Doubly Fed Induction Generator Model for Transient Stability Analysis. *IEEE Trans. On Energy Conversion* **2005**, *20*, 388–397, doi:[10.1109/TEC.2005.845523](https://doi.org/10.1109/TEC.2005.845523).
46. Slootweg, J.G.; Polinder, H.; Kling, W.L. Representing Wind Turbine Electrical Generating Systems in Fundamental Frequency Simulations. *IEEE Trans. On Energy Conversion* **2003**, *18*, 516–524, doi:[10.1109/TEC.2003.816593](https://doi.org/10.1109/TEC.2003.816593).
47. Chee-Mun Ong, *Dynamic Simulation of Electric Machinery: Using MATLAB/SIMULINK*. Prentice Hall: Upper Saddle River, N.J, USA, 1998.
48. Song, Z.; Xia, C.; Shi, T. Assessing Transient Response of DFIG Based Wind Turbines during Voltage Dips Regarding Main Flux Saturation and Rotor Deep-Bar Effect. *Applied Energy* **2010**, *87*, 3283–3293, doi:[10.1016/j.apenergy.2010.04.009](https://doi.org/10.1016/j.apenergy.2010.04.009).
49. Vijay Vittal and Raja Ayyanar, *Grid Integration and Dynamic Impact of Wind Energy*. Springer: New York, USA, 2013.
50. Xiaoyu Wang. Investigation of positive feedback anti-islanding scheme for inverter-based distributed generation, Ph.D. thesis, University of Alberta, Alberta, Canada, Jun 2008.
51. L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” J. Artif. Intell. Res., Jan.–June 1996; vol. 4, pp. 237–285.
52. C. J. C. H. Watkins and P. Dayan, Q-learning: Machine Learning., Aug. 1992; vol. 8,no. 3, pp. 279–292.
53. S. Fujimoto, H. van Hoof, and D. Meger Addressing function approximation error in actor critic methods. Presented at the ICML,2018. doi: <https://arxiv.org/abs/1802.09477>.
54. Yu, T.; Zhen, W.-G. A Reinforcement Learning Approach to Power System Stabilizer. In Proceedings of the 2009 IEEE Power & Energy Society General Meeting; IEEE: Calgary, Canada, July 2009; pp. 1–5.
55. Varma, R.K.; Siavashi, E.M. Enhancement of Solar Farm Connectivity With Smart PV Inverter PV-STATCOM. *IEEE Trans. Sustain. Energy* **2019**, *10*, 1161–1171, doi:[10.1109/TSTE.2018.2862405](https://doi.org/10.1109/TSTE.2018.2862405).