ENG720: Research Proposal

Title: Automatic generation control of a two area power system using deep reinforcement learning

Author: Shane Reynolds

Supervisor: Charles Yeo & TBC

Degree: Bachelor of Engineering (Honours)

Contents

1 Introduction & Background

In 2018 there was approximately 261TW h of power generation in the Australian electricity sector. Renewables contributed to 19% of the total generation, increasing from 15% in 2017 [1]. This is a pattern which has followed on from 2016. Indeed, a trend of greater renewable energy penetration in the electricity generation market has emerged starting in approximately 2008, as shown in Figure 1.

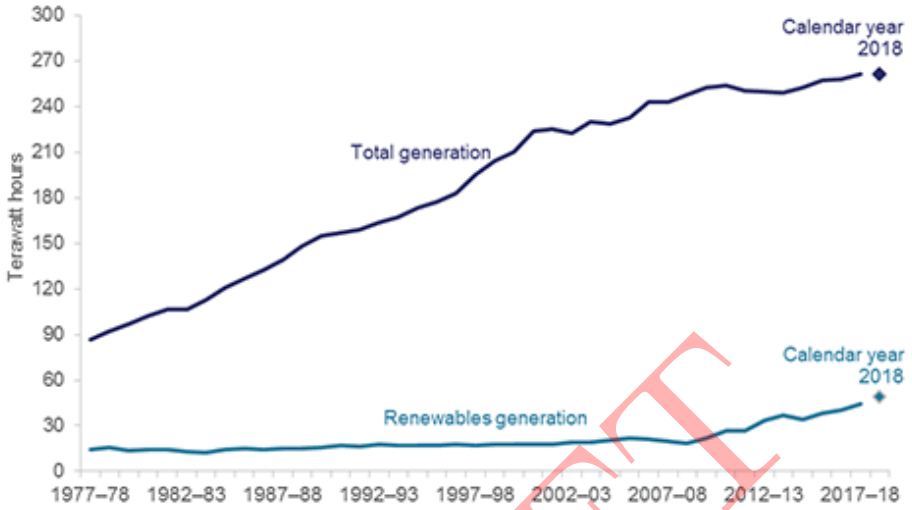


Figure 1: text

One of the benefits of transitioning from thermal sources of generation to renewable sources is reduced greenhouse gas emissions (REFERENCE), however, this transition is not without its drawbacks. Increased reliance on renewable generation sources poses some challenges regarding power system stability. A recent example is the system failure resulting from an event cascade, triggered by cloud cover shadowing a solar array in Alice Springs. The system failure resulted in a blackout occurring in Alice Springs for approximately 8 hours (REFERENCE). The Entura report highlighted that poor control policies were one of the many factors contributing to the blackout. In this instance, a generator provisioned to ramp up in the event of cloud cover limiting solar array output was unable to be controlled (REFERENCE). Moreover, generators that were still under the control regime were issued operating set points above their rated capacity, which eventually resulted in thermal overload and subsequent tripping from the protection system.

One of the issues with currently employed control methods (using classical techniques) is that they can be brittle when faced with system changes, or scenarios which they were not designed for/ envisioned. A more robust controller would be one that can learn and adapt to a system, given some broad control objective. (NEED TO REPHRASE THIS ASPECT). This research proposes a Deep Reinforcement Learning agent for controlling the frequency of a power system with multiple generators, and multiple stochastic loads. Existing approaches currently employ classical engineering control methodologies.

1.1 Power Systems and Frequency

Interconnected power systems are comprised of power generating units and energy storage systems connected to transmission and distribution networks. Generated power is used to service load demand. A single line diagram of a power network can be seen in Figure 1. The left hand side of the diagram shows thermal generation units, such as coal and nuclear, in addition to renewable sources of generation, like wind and solar. The right hand side of the figure shows the distribution network and the consumers of generated energy: industry and households.

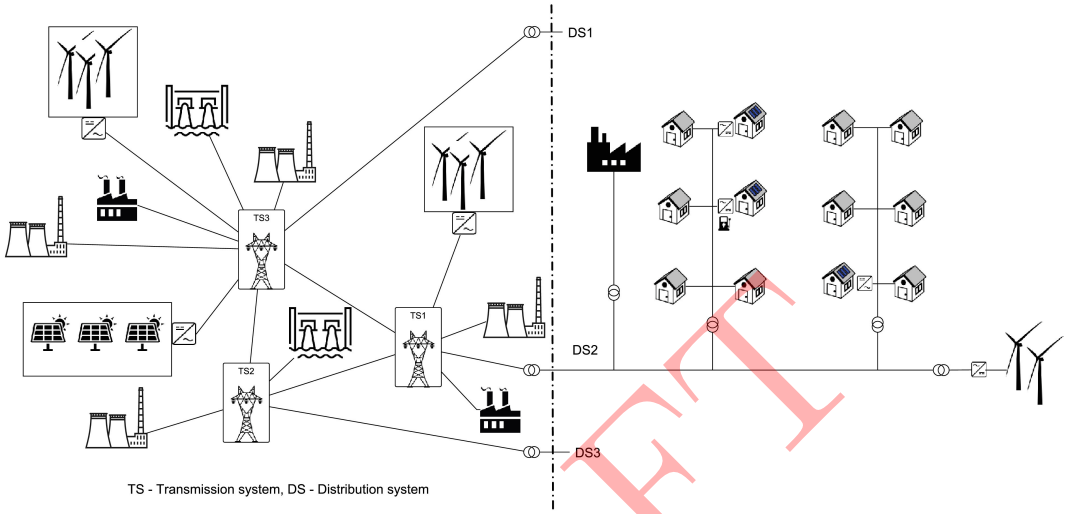


Figure 2: A single line diagram of a typical power system taken from [2]. The image shows points of generation from thermal and renewable sources, and the subsequent supply of generated energy to meet load demand through the transmission and distribution network.

One of the key elements to successful operation of interconnected power systems is ensuring total load demand is matched with total generation, taking into account power losses involved with generation, transmission, and distribution [3]. To understand why it is important to match generation with load demand it is useful to first consider the basic operation of a single thermal generator.

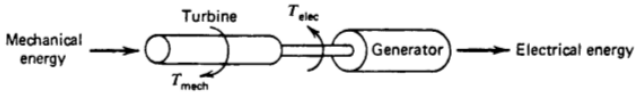


Figure 3: A thermal generation unit consists of a prime mover (turbine), and a synchronous machine. This image was taken from [3].

The essential elements of a thermal generator are a prime mover (turbine) and a synchronous machine, as depicted in Figure 2. The prime mover provides mechanical torque, T_{mech}, which drives the synchronous machine producing electrical energy. In response, the synchronous machine creates a torque which opposes T_{mech}, dependent on the size of the load demand from households and industry. This is referred to as

electrical torque and is denoted as T_{elec}. If α represents angular acceleration of the generator rotating mass, and I is its moment of inertia, then by Newton's second law:

$$\sum T_i = I\alpha \tag{1}$$

Equation (1) shows that when T_{mech} equals T_{elec} the system will be in a steady state with zero angular acceleration, and a constant rotation at some angular velocity ω. Now, if $T_{mech} > T_{elec}$, then the angular velocity ω of the system will increase as the system speeds up, resulting in a frequency increase in the system. Conversely, if $T_{mech} < T_{elec}$ then the angular velocity ω will decrease as the system slows down, resulting in a frequency decrease. What makes this situation interesting is that at any point in time the total electrical load demand will fluctuate stochastically, meaning that an uncontrolled system will have a continually changing frequency. Australia's electricity network is designed to operate at a frequency of 50Hz. In the majority of network scenarios AEMO has a desired operating range for frequency which lies between 49.85Hz and 50.15Hz [4]. Similarly, the PWC Network Technical Code for the Northern Territory states that under normal operating conditions frequency should be maintained in the range 49.80Hz to 50.20Hz [5]. Operation outside of specified ranges can cause damage to electrical equipment such as transformers or motors, which are designed to operate at specific frequencies [6]. Network designers engineer protection schemes so that sustained frequency excursions outside of the allowed range will cause equipment to trip from the network [7]. Protection schemes tripping equipment from the network is undesirable since this can leave households and industry without power, resulting in economic loss. Further, if disconnections are uncontrolled then this can lead to a further loss of system stability [7].

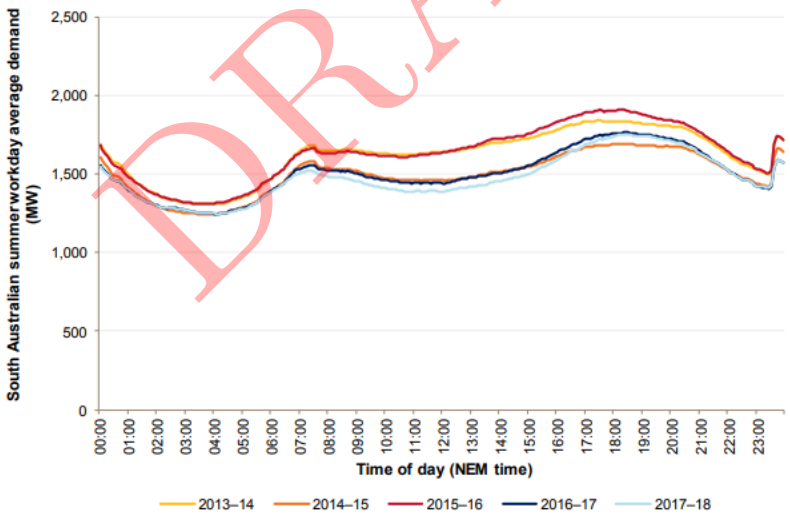


Figure 4: Weekday energy demand profile in South Australia during summer [8].

System controllers, such as the Australian Energy Market Operator (AEMO) and Power and Water Corporation (PWC), are therefore interested in being able to control the system to follow changes in load demand so that system frequency is maintained in the allowable range. Additionally, they are interested in control mechanisms to restore frequency excursions as a result of unexpected disturbances.

System controllers can use historical data to forecast daily demand profiles with some reliability. A plot of average historical data, for the daily demand profile in South Australia during Summer, can be seen in Figure 3. This type of forecasting does not help when trying to predict the occurrence of random disturbances, however, it does provide a starting point for estimating required generation needed to meet demand. It is important to note that forecasting is not perfect. Inevitably mismatches in supply and demand will occur causing small imbalances between T_{mech} and T_{elec}, resulting in a change to angular velocity ω and the network frequency [9]. To perfectly match supply and demand, system controllers use generators referred to as regulating units [10]. A regulating unit is a generator that has capacity to increase or decrease mechanical torque T_{mech} allowing the system controller to provide two functions: load following; and restoring the system to stable operating conditions in the event of a disturbance [11]. Using a regulating unit to load follow is referred to as the provision load following ancillary services [12]. Load following control adjusts regulating units slightly to match supply perfectly with a demand load profile, like that shown in Figure 3. Using a regulator to restore the system after a disturbance is referred to as providing spinning reserves [12]. When used in this fashion the regulating unit is not responsible for arresting frequency excursions, rather, it is used to restore the system back to the allowable frequency operating range after the frequency excursion has been arrested. An example of a frequency excursion, arrest, and subsequent restoration can be seen in Figure XXXX.

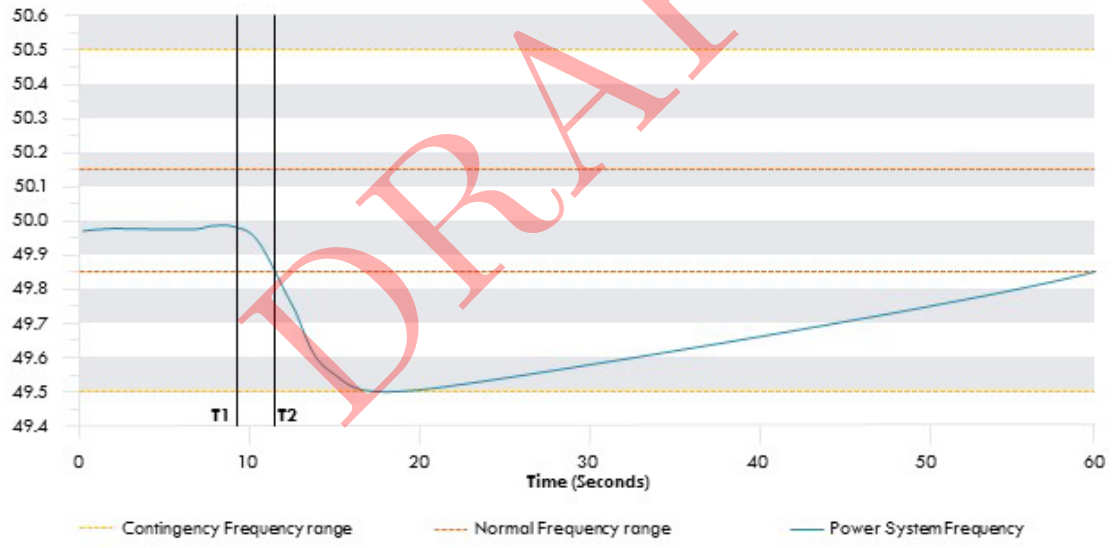


Figure 5: A frequency disturbance occurs just before the 10 second mark, and regulating units ramp up their generation to first arrest the disturbance, and provide the subsequent correction, returning system frequency to 50Hz.

AEMO and PWC do not require all generators on the network to act as regulating units since adequate frequency control can be achieved using a subset of the available generators.

1.2 Frequency control for a single area system

The power system model shown in Figure 1, on page 2, depicts total generation coming from many generation assets - this is complex to model. Researchers often find it useful to divide generation assets into sub-groups referred to as control areas [10]. A control area is defined as a subset of generators which are in close proximity to each other and constitute a coherent group that speed up and slow down together, maintaining their relative power angles [10]. The total network is therefore comprised of many interconnected control areas. An example of a series of interconnected control areas can be seen in Figure 5. Notice that for each area there is only a single load and a single generator. Typically, for each control area, researchers will aggregate many loads into a single load, and many generators into a single generator. This simplifies the model further [11].

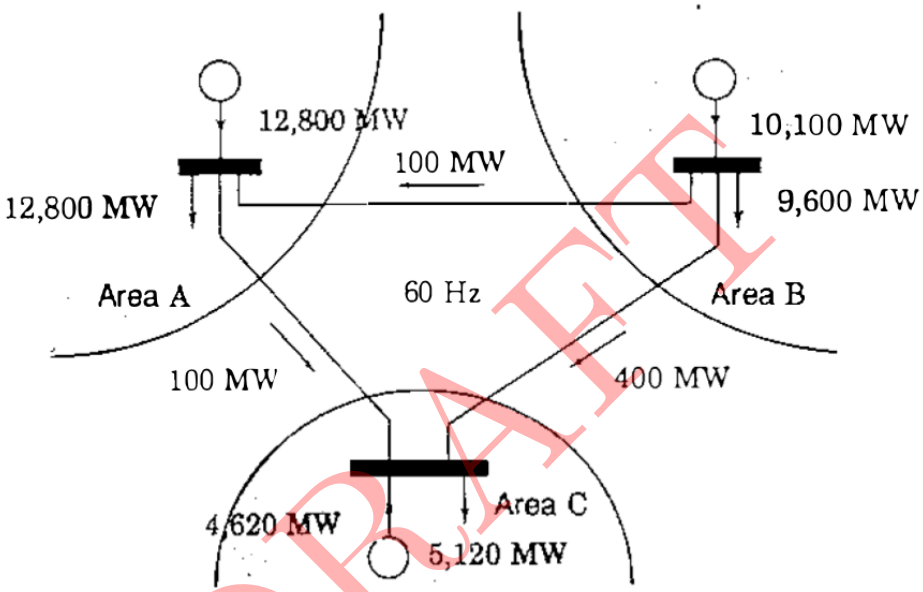


Figure 6: An example of three interconnected control areas in a 60Hz power system. The interconnections allow power to flow from one area to another, allowing generators to service loads from different areas. Each control area is consists of many generators and loads, but are modelled with a single generator and single load, respectively [11].

The simplest power system to control is one that consists of a single control area, such as Area A in Figure 5. This power system has no interconnections to any other control area. It is comprised of consumer load demand, and a set of generators, some of which are acting as regulating units. As previously mentioned, for modelling simplicity, loads are aggregated to a single load, and generators are aggregated to a single generator. This classic, simple system is well understood and it is generally acknowledged that a governor feedback control regime can successfully perform frequency control of the system [3, 11, 10]. Most introductory textbooks on power systems cover governor control of this system. Kothari (2011) provides a particularly well laid out approach to developing linear models for the turbine,

generator load, and governor for the system - the full model can be seen in Figure 6 [10]. The leftmost block is a first-order linear model of the speed governor, the second block is a first-order model of the turbine, which the controller directly acts on. The final block is the generator load, which is also a first order system. The over all system model is a second order linear model, with a first order controller.

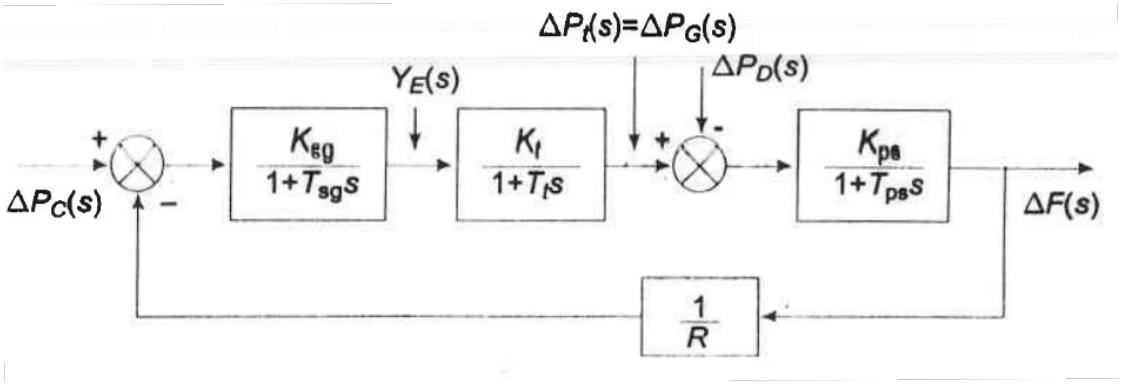


Figure 7: A classical feed back control approach to a second order linear system. The second order system is comprised of a first order model for both the turbine, and generator load. The controller is modelled as a first order system [10].

1.3 Frequency control for two area system

The system presented in Section 2.2 is useful to help understand the role of governors in controlling power system frequency, however, a single area model is too simple. In reality, power systems are comprised of many control areas connected through tie lines, which are typically transmission lines [13]. Often it is the case that there is some net power transfer over the tie lines, enforceable by contract. Distinct control areas are typically thought of as different participants in the generation market, or simply as different regions in which generation assets are based [13]. The simplest model which includes tie lines is the two area power system [10].

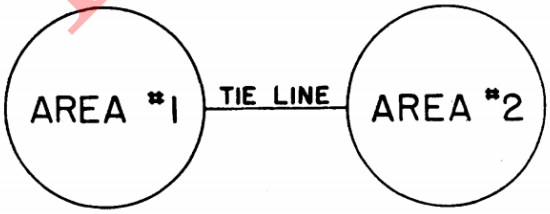


Figure 8: Two area power system is comprised of generators and load connected via a tie line. Power flows from one area to the other depending on economic contracts.

The control objective now is to maintain the inter-area power transfer, whilst regulating the frequency of each area. Simply relying on governor control in each area will not satisfy the control objective. Suppose control area 1 was supplying a 50MW load, and was contracted to supply control area 2 with 20MW. If control area 2 also has a 50MW load, then it is supplying only 30MW to satisfy the demand in this

area. Now, consider a 30MW load increase in the demand for area 1. Relying on governor control will see generators from both area 1 and area 2 speed up in response to this step change. Ultimately, the increased power demands will be met, however, the power transfer over the tie line is likely to be less than the contracted 20MW value - this is problematic. Contract violations due to system instability and control issues do not allow for a stable market in which energy can be reliably traded. Fortunately, two area power systems are well understood. Linear models have been developed to simulate these systems, and classical control approaches have been successfully implemented to meet the new control objectives. In order to achieve this, a metric called Area Control Error (ACE) is used. This measures the distance from a target frequency as well as the deviation from tie-line contractual obligations. The implementation of this control system is shown in Figure XXXX.

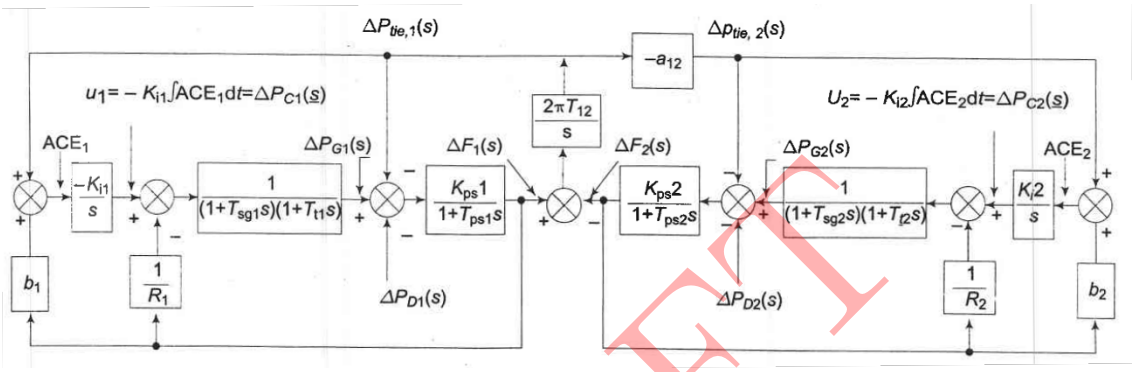


Figure 9: A classical feed back control approach to a two power area system [10].

1.4 Reinforcement learning

Reinforcement Learning (RL) is a branch of machine learning that is concerned with how agents make sequential decisions to maximise some notion of a cumulative reward. It is a simple idea, but one which can be generalised to applications in many different fields, such as NOTE FIELDS HERE. Subsections 1.4.1 and 1.4.2 provide a brief overview of key architectural components of RL, and subsection 1.4.3 gives details on how these components are implemented to build an agent that can perform a control activity.

1.4.1 Markov decision process

Suppose that a robotic agent exists in some environment which is comprised of many discrete states, $s \in S$, such that S denotes the state space. At any discrete point in time the agent can take an action $a \in A$, where A denotes the action space. When the agent takes an action in a given state, the agent receives some reward, denoted with $r \in R$, where R is the reward set. If an agent is in a given state, s, and takes and action, a, this will transition the agent to a new state, s', and yield reward, r, with some given probability - these are referred to as state transition probabilities.

Transition probabilities are denoted as follows:

$$P(S_{t+1} = s', R_{t+1} = r \mid S_t = s, A_t = a) \tag{2}$$

The set of parameters, outlined above, make up a framework referred to as a Markov Decision Process (MDP) [14].

1.4.2 Return, episodes, and policy

As the robotic agent takes actions at each discrete time step, it receives a reward. The cumulative sum of this reward is referred to as the return [14]. The return is denoted, for N discrete time steps, as:

$$G_t = r_t + r_{1+1} + r_{t+2} + \ldots + r_{N-1} \tag{3}$$

Often it is convenient to make future rewards less important than more immediate rewards. This is achieved by multiplying each reward in the sequence by a discount factor, $\gamma \in [0, 1]$. Equation (XXXX) then becomes:

$$G_t = r_t + \gamma r_{1+1} + \gamma^2 r_{t+2} + \ldots + \gamma^{N-1} r_{N-1} = \sum_{k=0}^{N-1} \gamma^k r_{t+k} \tag{4}$$

Typically the duration of time that an agent will cumulate reward over is referred to as an episode. An episode is made up of a beginning, middle, and an end. Typically this consists of an RL agent beginning in some initial state, a period of time passes where the agent takes actions and undergoes state transitions, and then the episode concludes when the agent reaches a terminal state. At the episode conclusion, the agent receives it's cumulative reward.

Finally, in order for the robot to act within the environment, it needs to have a policy. A policy, π, is defined as a mapping from states to actions, that is, a rule which determines what action the robot will take for a given state. A deterministic policy, $\pi(s)$, maps a single action to a single state. A stochastic policy, $\pi(a|s)$, defines a probability distribution over the actions for a given state. An optimal policy, denoted π^*, is a policy which will maximise the cumulative reward that the agent receives over an episode.

1.4.3 How does an RL agent learn?

The main objective of RL is to develop an optimal policy. There are many algorithmic approaches to building an optimal policy, but most focus updating the value for a state action pair as the agent takes actions in different states. The agent normally starts with a randomised policy meaning that it will take actions at random as it explores the state-action space. As values are assigned to state-action pairs during an episode, the agent modifies the policy. This process is repeated for many episodes

Q-table initialised at zero

	UP	DOWN	LEFT	RIGHT
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0

After few episodes

	UP	DOWN	LEFT	RIGHT
0	0	0	0	0
1	0	0	0	0
2	0	2.25	2.25	0
3	0	0	5	0
4	0	0	0	0
5	0	0	0	0
6	0	5	0	0
7	0	0	2.25	0
8	0	0	0	0

Eventually

	UP	DOWN	LEFT	RIGHT
0	0	0	0.45	0
1	0	1.01	0	0
2	0	2.25	2.25	0
3	0	0	5	0
4	0	0	0	0
5	0	0	0	0
6	0	5	0	0
7	0	0	2.25	0
8	0	0	0	0

Figure 10: The Q-table on the left shows the initialised policy when the agent begins learning. The middle and rightmost Q-tables show the agent developing an understanding of which actions are valuable in which states.[15]

and eventually the agent policy converges to an optimal policy. The literature often presents policy values in a table. An example of this can be seen in Figure XXXX. The rows represent the different states, and the columns represent different actions. The values in each cell provide an ordinance on how valuable each action is for a given state. These types of tables are referred to as a Q-tables. One of the benefits learning like this is that it is only necessary for the agent to understand inputs that uniquely define a state, and what actions are available to it in each state. It is not necessary for the agent to know the state transition dynamics of the system this means that an agent can learn to control a system for which is does not have a mathematical model.

1.5 Deep reinforcement learning

For low dimensional state-action spaces RL approaches result in reasonable control performance, however, as state space dimensionality increases models like Q-tables experience difficulty. The main reason for this is simply that it becomes difficult for the discrete RL algorithm to visit every state action pair resulting in unchanged values for an increasing number of state-action pairs. Essentially this means that the agent does not have complete knowledge of optimal actions for every given state, leading to the derivation of sub-optimal policies. To get around this problem, for RL problems with high dimensional state spaces, the discrete Q-table is replaced with a function approximator known as a neural network. A high level overview of the architecture can be seen in Figure 11. It is the neural network architecture in which the agent policy is implemented. As the agent learns, it adjusts weights in the neural network to change the policy. This is approach is powerful because neural

networks are good at generalising, and hence the agent does not need to visit every state action pair to be able to make good decisions.

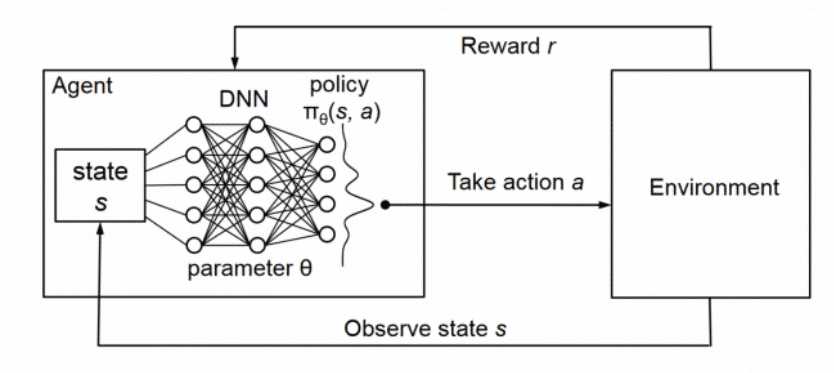


Figure 11: The agent interacts with the environment by taking actions, which affect the state it is in and the reward it receives. The rewards allow agent to adjust the weights in the neural net to build better policies.

2 Research Aims

The principle aim of this research is to compare the performance of classical engineering control methodologies against a control system based on a DRL agent, when tasked with performing AGC for load following tasks on a two area power system. The ultimate research outcome will be to provide comment on the feasibility of using DRL agents for this purpose.

3 Scope

The proposed research focuses DRL agent performance against the task of load following by maintaining system frequency within PWC's allowable region of 49.80Hz to 50.20Hz for normal operating conditions. Time permitting, agent performance against the task of frequency restoration following a disturbance event may be considered. The key performance aspects that will assessed are the controller's ability to:

- maintain system frequency to the desired nominal 50Hz value;
- maintain the tie line power flow between control areas at the scheduled value.

The research will not consider power systems larger than a two area power system, nor will it consider the control of system variables other than frequency. Note, however, other system variables may be used as input features under agent training and inference regimes. Comparison of DRL agent performance will be made against theoretical models of classical control architectures. Performance against

control architectures implemented in practice will not be considered. Research will be conducted in an entirely simulated environment, and agent performance on real hardware will not be explored.

4 Approach

4.1 Required data sources and data management

Training a DRL agent to change regulating generator set points in order to maintain system frequency and tie line contractual obligations while load following will require realistic demand profiles. Similarly, performing system restoration after a disturbance will require realistic disturbance scenarios. Ideally this data would come from a major utility, such as PWC, in the form of a time series dataset with a high number of features, and short durations between each observation. Data acquisition will be one of the principle objectives in the early stages of research. Should the acquisition of data from PWC or TGEN be viable, a data management plan will need to be developed which addresses concerns around the sensitivity and security of the data, where it will be stored, and data treatment (or disposal) once the research is concluded.

In the event data cannot be acquired from a utility, a synthetic data set may need to be derived. This would be achieved by understanding key statistical parameters of a typical load demand profile, and using these to create a stochastic process which emulates the load demand signal. This could also be done for other system variables, however, care would need to be taken ensuring correlations are preserved between multiple variables in the synthetic time series dataset.

4.2 Theoretical approach

In order to establish the most effective way to approach this research problem, a clear understanding of the benefits and limitations of existing AGC approaches is needed. Determining justifications for practical AGC design choices will help to uncover important performance aspects the research should focus on. Equally important is exploring alternative approaches to AGC that researchers have investigated historically. This should have a particular focus on the use of Neural Networks and DRL agents for AGC. A literature review will be the main avenue for achieving this.

As discussed in Section 4.1, securing load demand profile datasets from a major utility, or developing synthetic load profile datasets based on local load profile characteristics is an important aspect of the research and will need to be conducted as early as possible. Similarly, investigating suitable software packages (open source or commercial) to develop the power system simulation model, and investigating suitable programming languages to implement DRL agent should be explored. This information will most likely be found when exploring the field literature. It will be

important to understand how other researchers integrated the DRL agent implementation with the simulation environment.

A simulated model of the two area power system will be developed. The decision to use a linear or non-linear model will be informed by the literature review. It may be interesting to explore DRL agent performance on both linear and non-linear models since one of their advantages is that they have a demonstrated capacity for controlling highly non-linear systems. Classical engineering system modelling techniques, like those seen in Ogata, will be employed for power system model development (REFERENCE). An area of interest in this domain is how sensitive a DRL agent control regime is to changes in key plant parameters - for a given set of parameter changes both DRL agent and classical control architecture performance could be compared to see which controller is more brittle.

A feedback loop controller will be developed for the two area power system using models presented in literature such as Kothari [10]. This is an application of classical engineering control theory taught in most undergraduate engineering courses. A DRL agent will be developed using an architecture that provides for continuous input signals and continuous output signals. There are a number of well established DRL models presented in texts like Sutton and Barto that will be explored to determine the most appropriate approach. Time permitting, a DRL model that uses discretised input and output signals will be developed. Discrete models offer lower performance due to errors introduced in the discretisation process, but can be computationally less expensive than continuous models. DRL models will be trained using data previously acquired either from the a utility, or by synthesis. Metrics will be selected to measure the performance of both controllers. Choice of metrics will be informed by earlier research (mentioned above). Control models will be compared, and differences in their performance will be compared - this will be one of the major focuses of the research.

The full list of tasks for the research design are as follows:

1. Enquire with power utility to secure data
2. Investigate ways to synthesise data
3. Develop data management plan
4. Literature review:
   (a) Benefits and limitations of existing AGC approach
   (b) Justification for design choices for practical AGC implementations
   (c) Performance measurement criteria for AGC
   (d) Alternate approaches to AGC (historical)
5. Investigate suitable software package to conduct simulation

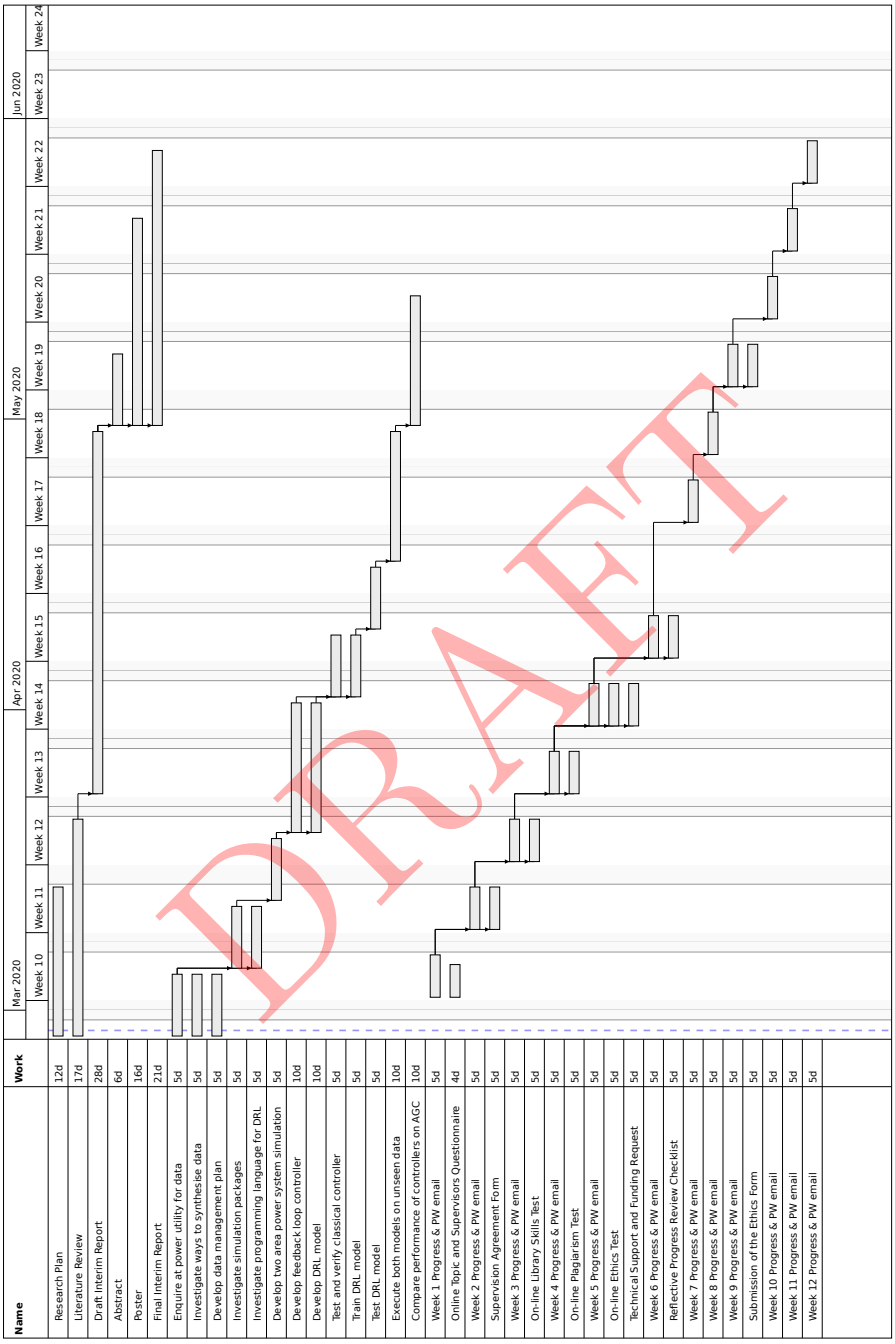
6. Investigate suitable programming language to implement DRL agent and integrate with simulation
7. Develop and test simulation of two area power system
8. Develop feedback loop controller for two area power system
9. Test classical controller
10. Develop DRL model
11. Train and test DRL model
12. Execute control trials on both models for an unseen sequences of load demand data
13. Compare controller performance on AGC task

It is anticipated that there may be some issues in carrying out the aforementioned research design. The biggest risk would the inability to successfully build the control models for both the classical engineering controller, and the DRL controller. For the classical engineering controller, the issue would be present as the inability to find the appropriate parameter settings to deliver stable control. With the DRL controller, the problem is selection of an appropriately sized neural network, and training hyper-parameters.

5 Deliverables Specification

The main deliverable from this research is a conclusion about the feasibility of using DRL agents to provide AGC for a two area power system. The expectation is that the DRL agent will perform at least as well as a classical control approach. This expectation is based on the existing research in this field, which has shown that standard RL agents can perform AGC at least as well as classical control methods (REFERENCE). It is envisaged that should the research meet expectations, this could provide a pathway for investigation of more exotic DRL models to find better power system controllers. The ultimate goal in this field would be to have a DRL agent that is always learning, so that it can adapt control strategies to unseen system conditions providing a more flexible and less brittle controller. This would help to avoid problems like that seen in the Alice Springs system black event.

6 Timeline



7 Resources

The main research objectives, at a minimum, will require a computer with a Linux operating system, RAM, and a Graphics Processing Unit (GPU). The exact specifications of these computational resources are not yet known, and will be dependent on the size of the datasets, and the amount of computation required to optimise the the DRL agent's cost function. Currently, access has been provided to an Intel Quad Core 3.2GHz i5 CPU machine with 8GB of RAM, Nvidia GTX960 GPU, and Ubuntu operating system. Most likely these resources will suffice, however, in the event that greater computational power is needed, access to a high end virtualised Amazon Web Services (AWS) system set up to conduct DRL research will be required. Currently, such an instance has been provisioned on an existing AWS account and is currently deactivated to avoid cost. Using such a system can be expensive (1 hour of compute is approximately 20AUD), however, AWS will often supply student research with monetary credit.

Table 1: text

References

[1] Department of industry science energy and resources. Electricity Generation. https://www.energy.gov.au/data/electricity-generation, 2020.

[2] M. Glavic. (Deep) Reinforcement learning for electric power system control and related problems: A short review and perspectives. *Annual Reviews in Control*, 2019.

[3] A. J. Wood, B. F. Wollenberg, and G. B. Sheble. *Power generation, operation, and control*. Wiley, 3 edition, 2013.

[4] Power system frequency and time deviation. Technical report, Electricity system operations planning and performance, July 2012.

[5] Power and Water Corporation. *Network technical code and network planning criteria*, 4 edition, December 2018.

[6] P. C. Sen. *Principles of Electric Machines and Power Electronics*. Wiley, 2014.

[7] Power system frequency risk review - draft report. Technical report, Electricity system operations planning and performance, April 2018.

[8] South Australian Electricity Report. Technical report, Australian Energy Market Operator, November 2018.

[9] J. D. Glover, S. S. Mulukutla, and T. J. Overbye. *Power system analysis and design*. Cengage Learning, 5 edition, 2012.

[10] D. P. Kothari and I. J. Nagrath. *Modern Power System Analysis*. McGraw Hill India, 4 edition, 2011.

[11] J. J. Grainger and W. D. Stevenson. *Power System Analysis*. McGraw Hill, 1994.

[12] Australian Energy Market Operator. Ancilliary services. https://aemo.com.au/en/energy-systems/electricity/wholesale-electricity-market-wem/system-operations/ancillary-services, 2020.

[13] NPTEL. Economic Operation of Power Systems. https://nptel.ac.in/content/storage2/courses/108104051/chapter_5/5_7.html, 2020.

[14] R. S. Sutton and A. G. Barto. *Reinforcement Learning*. MIT Press, 2018.

[15] S. Gite. Practical reinforcement learning: getting started with q-learning. https://towardsdatascience.com/practical-reinforcement-learning-02-getting-started-with-q-learning-582f63e4acd9, 2020.