

REINFORCEMENT LEARNING FOR SMART GRID OPTIMIZATION

Thesis submitted to the
Indian Institute of Technology Kharagpur
for the award of the degree

of

Doctor of Philosophy

by

Sourav Chatterjee
(Roll No. 21EE92R07)

Under the supervision of
Prof. Arindam Sen
Prof. Meera Iyer

Department of Electrical Engineering
Indian Institute of Technology Kharagpur
July 2026

CERTIFICATE

This is to certify that the thesis entitled **Reinforcement Learning for Smart Grid Optimization**, submitted by Sourav Chatterjee to the Indian Institute of Technology Kharagpur, is a record of bona fide research work carried out under our supervision and is worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute.

Prof. Arindam Sen

Department of Electrical Engineering

Prof. Meera Iyer

Department of Electrical Engineering

DECLARATION

I certify that the work contained in this thesis is original and has been done by me under the guidance of my supervisors, that it has not been submitted to any other institute for any degree, and that whenever I have used materials from other sources I have given due credit to them in the text and listed them in the references.

Sourav Chatterjee

Acknowledgements

I express my gratitude to my supervisors, Prof. Arindam Sen and Prof. Meera Iyer, for their guidance and encouragement. I thank the members of my Doctoral Scrutiny Committee for their suggestions, and my colleagues in the power systems laboratory for their support.

Abstract

Distribution grids with high shares of rooftop solar must balance generation and demand in real time, and conventional optimization struggles when forecasts are uncertain. This thesis applies reinforcement learning to two grid-control problems: volt-var control with smart inverters and scheduling of community battery storage. It formulates both as Markov decision processes, trains policies with a safety layer that projects actions onto the feasible set of the power flow constraints, and evaluates them on a modified 33-bus test feeder. The learned volt-var policy keeps all bus voltages within limits in 99.6% of simulated hours while reducing network losses by 11% compared with a droop-control baseline.

Keywords: reinforcement learning, smart grid, volt-var control, battery scheduling, safe learning

Contents

Acknowledgements	iii
Abstract	iv
1 Introduction	1
2 Background	2
3 Safe Reinforcement Learning for Volt-Var Control	3
4 Results	4
5 Conclusion	5
A Test Feeder Modifications	6

List of Figures

3.1	The actor's proposal is projected onto the safe set before it reaches the inverters.	3
-----	---	---

List of Tables

4.1	Volt-var control over one simulated year.	4
-----	---	---

Chapter 1

Introduction

Rooftop solar has turned many distribution feeders from one-way conduits into two-way networks. Midday generation can push voltages above their limits, while evening demand pulls them below. Utilities regulate voltage with tap changers and capacitor banks designed for slower, more predictable variation. Smart inverters and batteries respond in seconds, but deciding how they should respond, under uncertain forecasts, is a hard sequential decision problem.

Chapter 2

Background

For a radial feeder, the linearized DistFlow model relates voltage magnitudes to active and reactive power injections [1]. Volt-var control chooses inverter reactive power q_t to keep voltages v_t within $[\underline{v}, \bar{v}]$ while minimizing losses. We treat it as a Markov decision process and learn a policy with an actor-critic method [3], maximizing

$$J = \mathbb{E} \left[\sum_t \gamma^t (-P_{\text{loss},t} - \lambda \| \max(0, v_t - \bar{v}) + \max(0, \underline{v} - v_t) \|_1) \right]. \quad (2.1)$$

Chapter 3

Safe Reinforcement Learning for Volt-Var Control

A penalty in the reward discourages violations but does not prevent them. We add a safety layer that projects each proposed action onto the set of actions predicted by the linearized power flow to keep all voltages within limits [2]. The projection is a small quadratic program solved at every step.

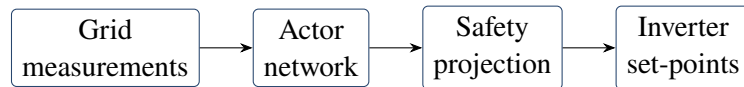


Figure 3.1: The actor’s proposal is projected onto the safe set before it reaches the inverters.

Chapter 4

Results

Table 4.1 compares controllers over one simulated year on the modified 33-bus feeder.

Table 4.1: Volt-var control over one simulated year.

Controller	Hours within limits (%)	Losses (MWh)
Droop control	97.8	212
Reinforcement learning	98.1	186
Safe reinforcement learning	99.6	189

Chapter 5

Conclusion

Reinforcement learning can coordinate fast grid resources better than fixed rules, and a projection onto the safe set keeps it within operating limits. Field validation on a live feeder is the next step.

Appendix A

Test Feeder Modifications

Photovoltaic systems of 150 to 400 kW were added at eight buses of the 33-bus feeder, together with two 500 kWh community batteries at buses 18 and 33.

Bibliography

- [1] Mesut E. Baran and Felix F. Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, 4(2): 1401–1407, 1989.
- [2] Gal Dalal, Krishnamurthy Dvijotham, Matej Vecerik, Todd Hester, Cosmin Paduraru, and Yuval Tassa. Safe exploration in continuous action spaces. arXiv:1801.08757, 2018.
- [3] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In *International Conference on Learning Representations*, 2016.