# Learning Decentralized Frequency Controllers for Energy Storage Systems

Zexin Sun* Zhenyi Yuan* Changhong Zhao Jorge Cortés

***Abstract*— This paper designs decentralized controllers for energy storage systems (ESSs) to provide active power control for frequency regulation. We propose a novel safety filter design to gracefully enforce the satisfaction of the limits on the state of charge during transients. Our technical analysis identifies conditions on the proposed design that guarantee the asymptotic stability of the closed-loop system with respect to the desired equilibria. We leverage these results to provide a controller parameterization in terms of a single-hidden-layer neural network that automatically satisfies the conditions. We then employ a reinforcement learning approach to train the controller to optimize transient performance in terms of the maximum frequency deviation and the control cost. Simulations in an IEEE 39-bus network validate the significant transient performance improvements of the proposed controller design.**

## I. Introduction

With the increasing penetration of renewable energy sources such as wind turbines and photovoltaics in power systems, the frequency may fluctuate rapidly in an extensive range, which can lead to abnormal operation of electrical appliances and increase the risk of large-scale power outages. The intrinsic volatility and uncertainty of renewable generation require sufficient backup resources, which motivates the proliferation of energy storage systems (ESSs) in modern power grids [1]. This paper is motivated by the aim of facilitating the use of ESSs to maintain frequency stability and enhance transient performance after disturbances.

*Literature Review:* A lot of recent attention has been devoted to optimal frequency control, which combines frequency control with optimal dispatch of power systems to achieve simultaneously stable and economic operation [2]–[5]. In this context, distributed load frequency control has been widely investigated [6]–[10]. In [6], a primary frequency control problem is formulated and optimal decentralized controllers for responsive loads are derived from a partial primal-dual gradient algorithm. This framework is extended in [7] to secondary frequency control to recover the nominal frequency accounting for line congestion. Further extensions for optimal distributed frequency control include nonlinear power flow and higher-order generator models [8], variations in renewable generation [9], and nonsmooth cost functions [10].

These studies focus on asymptotic stability and optimality of the steady state post-disturbance, but do not examine the safety and optimality of the transient. In recent research efforts, [11] proposes a reinforcement learning framework to develop primary frequency control with stability guarantees and transient performance enhancement. This idea is revisited in [12] by relaxing the stability conditions, incorporating safety considerations, and is further extended to secondary frequency control in [13] and adaptive control handling time-varying load with frequency restoration [14]. However, these works only consider controllable generators or loads with no restriction on the delivered/absorbed energy over time and are therefore not applicable to ESSs, hence not realizing their potential for frequency regulation.

*Statement of Contributions:* ESSs have the capability to improve transient frequency performance due to its fast speed of charging and discharging. However, its restricted state of charge (SOC), i.e., the limited amount of energy to be charged or discharged, adds a critical challenge to the controller design. This paper tackles this challenge by studying the participation of ESSs in frequency control. We make the following contributions. First, with a novel safety filter design, we propose a class of controllers that satisfy the SOC restriction of ESS automatically. We identify conditions on this controller design that ensure closed-loop stability. Then, we construct neural networks to parameterize the proposed controllers which satisfy the identified conditions by design. Leveraging a reinforcement learning framework based on recurrent neural networks, we train the controller to optimize the transient performance of frequency control after disturbances, specifically to reduce the maximum frequency deviation and the cost of control.

## II. System Model

Here, we introduce the dynamical model[1]. We model the power network as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ and $\mathcal{E}$ are the set of buses and lines, resp. Each line has an arbitrarily assigned reference direction and is denoted as an ordered pair, e.g., $(i, j) \in \mathcal{E}$ pointing from $i$ to $j$; the actual power flow may be opposite to the reference direction by taking a negative value. Without loss of generality, we assume each bus aggregates four power injections: a thermal generator, an uncontrolled net load that equals load power consumption minus undispatchable (renewable) generation, a fast responsive load or generator, and an energy storage system (ESS). As a basis of the dynamical model we will introduce, the power network is initially working at a steady

*Equal contribution. This work was supported by Hong Kong Research Grants Council through grant GRF 14212822 and NSF Award ECCS-1947050. Z. Sun is with the Division of Systems Engineering, Boston University (e-mail: zxsun@bu.edu). Z. Yuan and J. Cortés are with the Department of Mechanical and Aerospace Engineering, UC San Diego (e-mail: {z7yuan,cortes}@ucsd.edu). C. Zhao is with the Department of Information Engineering, The Chinese University of Hong Kong (e-mail: chzhao@ie.cuhk.edu.hk).

[1]We use $\mathbb{R}$ to denote the set of real numbers. We use $(\cdot)^\dagger$ and $(\cdot)^\top$ to represent the pseudo-inverse and transpose, resp. $\mathbf{1}$ and $\mathbf{0}$ are vectors of all ones and zeros of appropriate dimensions, resp. With a slight abuse of notation, we use $|\cdot|$ to denote the absolute value for real-valued arguments, and the cardinality when the argument is a set. A continuous function $\alpha : \mathbb{R} \to \mathbb{R}$ is of (extended) class-$\mathcal{K}$ if it is strictly increasing and $\alpha(0) = 0$, and is $\mathcal{K}_\infty$ if it is of class-$\mathcal{K}$ and $\lim_{r\to+\infty} \alpha(r) = +\infty$, $\lim_{r\to-\infty} \alpha(r) = -\infty$. Finally, $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and infinity norm, resp.

state:

$$\omega_i^0 = \omega^s, \tag{1a}$$

$$P_i^{m0} + P_i^{c0} - P_i^{l0} - \sum\nolimits_{(i,j)\in\mathcal{E}} B_{ij} \sin(\theta_i^0 - \theta_j^0) = 0, \tag{1b}$$

for all $i \in \mathcal{N}$, where $\omega^s$ is the nominal frequency, e.g., 50 or 60 Hz, and $\theta_i^0$ is the corresponding bus voltage phase angle. $P_i^{l0}$, $P_i^{c0}$, and $P_i^{m0}$ are resp. the initial uncontrollable net load, initial power output of the fast responsive resource, and initial thermal generator power output. The constant $B_{ij} > 0$ is the absolute value of susceptance of line $(i,j) \in \mathcal{E}$.

At the initial steady state: (1a) means that all buses are at the nominal frequency; (1b) means that the thermal generator power output $P_i^{m0}$ and the power output of fast responsive resource $P_i^{c0}$ are balanced with the uncontrolled net load $P_i^{l0}$ and the total power flow out of each bus $i$.

The state of charge (SOC) of ESS at each bus $i$ does not appear in (1). For convenience, we shift the SOC so that the feasible range is $[\underline{E}_i, \overline{E}_i]$, with $\underline{E}_i = -\overline{E}_i$. At the initial steady state, SOC $E_i^0 \in [\underline{E}_i, \overline{E}_i]$ and the ESS net charging power (which indicates discharge if it is negative) is $P_i^e = 0$.

We consider a change that occurs from the steady state above at time $t = 0$: the uncontrolled net loads $P_i^{l0}$ are disturbed by a step change $P_i^l$ at (a subset of) buses in $\mathcal{N}$, triggering power imbalance and frequency variation. According to the classic model of power network dynamics [15], and omitting the time index $t$ for brevity, we consider

$$\dot{\theta}_i = \omega_i, \tag{2a}$$

$$M_i\dot{\omega}_i = -D_i\omega_i - \sum_{(i,j)\in\mathcal{E}} B_{ij}\sin(\theta_i + \theta_i^0 - \theta_j - \theta_j^0) + \sum_{(i,j)\in\mathcal{E}} B_{ij}\sin(\theta_i^0 - \theta_j^0) + P_i^m + P_i^c - P_i^l - P_i^e, \tag{2b}$$

$$T_i^m \dot{P}_i^m = -P_i^m - \frac{\omega_i}{R_i}, \tag{2c}$$

$$\dot{P}_i^c = -\frac{\omega_i}{K_i}, \tag{2d}$$

$$\dot{E}_i = P_i^e, \tag{2e}$$

for all $i \in \mathcal{N}$, where $(\theta_i, \omega_i, P_i^m, P_i^c)$ are deviations from their initial states. In (2b), $M_i > 0$ and $D_i > 0$ are the inertial and damping constants, resp. Equation (2c) is a simplified model of generator dynamics, with time constant $T_i^m > 0$ characterizing the delay effect of the slowest component in the generator, typically the turbine. We assume the generator power setpoint does not change, and the change of $P_i^m$ is driven by a droop controller with factor $R_i > 0$. The change of power $P_i^c$ of the fast responsive resource is controlled by an integrator with $K_i > 0$ in (2d), which aims to achieve frequency restoration. Approximating the ESS charge/discharge efficiencies at $100\%$, the change $E_i$ in SOC can be calculated by integrating the charge/discharge power $P_i^e$ as $E_i(t) = E_i^0 + \int_0^t P_i^e(\tau)d\tau$, for all $i \in \mathcal{N}$ and all $t \geq 0$, which leads to (2e).

## III. Problem Formulation and Design

Here, we formulate the problem of interest and propose a novel controller design with safety filter to satisfy the SOC constraint of the ESS.

### A. Problem Formulation

We are interested in studying the participation of the ESS for frequency regulation of the power network (2), and particularly for enhancing its transient performance while guaranteeing closed-loop stability. Most existing works focus on the frequency regulation problem without the ESS, e.g. [6], [11] for primary frequency control and [7], [8], [13] for secondary frequency control. Specifically, we aim to design decentralized ESS controllers $\{P_i^e\}_{i\in\mathcal{N}}$ so that:

- System (2) converges to an equilibrium which has frequency at the nominal value;
- The ESS charge/discharge power satisfies:

$$\underline{P}_i^e \leq P_i^e(t) \leq \overline{P}_i^e, \ \forall i \in \mathcal{N}, \ \forall t \geq 0, \tag{3}$$

  where $\underline{P}_i^e < 0$ and $\overline{P}_i^e > 0$ are given rate limits for discharge and charge, resp.;
- The SOC of each ESS is restricted as:

$$\underline{E}_i \leq E_i(t) \leq \overline{E}_i, \ \forall i \in \mathcal{N}, \ \forall t \geq 0; \tag{4}$$

- The transient cost is minimized.

The control objectives above are formalized by:

$$\min_{\{P_i^e\}_{i\in\mathcal{N}}} \sum_{i\in\mathcal{N}} \Big( \|\omega_i\|_\infty + \frac{\rho}{T}\int_0^T g_i\left(P_i^e(t)\right) \mathrm{d}t \Big) \tag{5a}$$

$$\text{s.t. } (2), (3), (4) \tag{5b}$$

$$\lim_{t\to\infty} \omega_i(t) = 0, \forall i \in \mathcal{N}, \tag{5c}$$

where $T$ is the time horizon of interest and $g_i$ is a control cost function of the ESS operator's interest. The two terms in objective (5a) are respectively to minimize the maximum frequency deviation and the average accumulative cost of control during the transient. $\rho$ is a weighting factor to trade-off these two objectives.

The optimization problem (5) is non-convex and infinite-dimensional given that the decision variables are trajectories. Reinforcement learning (RL) offers an attractive solution approach to tackle this complexity. In RL [16], agents are free to interact with the environment and explore behaviors towards reward maximization to learn an optimal control policy. Existing RL approaches, e.g. [17], [18], neglect the hard constraints (3), (4) and the asymptotic constraint (5c), and instead consider soft penalization on these objectives. This does not guarantee system stability and safety, and might result in control policies that are not transferable to real implementations. Similar to recent works [11], [12], this motivates us to identify specific controller forms that meet these constraints by design, and then optimize the control policy using RL over this reduced class.

### B. Decentralized ESS Controller with Safety Filter

We propose the following decentralized ESS controller:

$$P_i^e = \mathcal{F}_i(\omega_i, E_i), \quad \forall i \in \mathcal{N}, \tag{6}$$

where $\mathcal{F}_i(\omega_i, E_i)$, $i \in \mathcal{N}$, is of the form:

$$\begin{cases} \max\left\{\frac{\underline{\alpha}_i(E_i - \underline{E}_i)}{\underline{\alpha}_i(\underline{E}_i^{\text{th}} - \underline{E}_i)} f_i(\omega_i), f_i(\omega_i)\right\}, & E_i \leq \underline{E}_i^{\text{th}}, \\ f_i(\omega_i), & \underline{E}_i^{\text{th}} < E_i < \overline{E}_i^{\text{th}}, \\ \min\left\{\frac{\overline{\alpha}_i(\overline{E}_i - E_i)}{\overline{\alpha}_i(\overline{E}_i - \overline{E}_i^{\text{th}})} f_i(\omega_i), f_i(\omega_i)\right\}, & E_i \geq \overline{E}_i^{\text{th}} \end{cases} \tag{7}$$

where $\underline{E}_i < \underline{E}_i^{\rm th} < \overline{E}_i^{\rm th} < \overline{E}_i$ and $f_i : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function that satisfies $\underline{P}_i^e \le f_i(\omega_i) \le \overline{P}_i^e$ for all $\omega_i \in \mathbb{R}$. A salient feature of our design (7) is the safety filter specified by $\mathcal{K}_\infty$ functions $\underline{\alpha}_i(\cdot), \overline{\alpha}_i(\cdot)$, activated at the thresholds $\underline{E}_i^{\rm th}, \overline{E}_i^{\rm th}$. The safety filter plays a critical role in preventing the violation of the SOC limits, as we discuss below. The controller (7) is a continuous function of $\omega_i \in \mathbb{R}$, particularly at the breaking points $\underline{E}_i^{\rm th}, \overline{E}_i^{\rm th}$.

We explain the rationale behind our design (6), (7). Most existing primary frequency controllers are functions of the frequency deviation $\omega_i$ only, assuming that the controlled resources can change their active power output as needed for stabilizing the frequency excursion. However, this is not the case here, because an ESS can only operate within its SOC limits, beyond which it cannot deliver or absorb active power anymore. This motivates us the addition of $E_i$ as another input of the control function $\mathcal{F}_i$ to decide the control action according to the SOC in real time.

The terms related to $E_i$, which we call *safety filter*, take effect only when needed and in a desired manner. For instance, in case that $E_i \le \underline{E}_i^{\rm th}$, i.e., the SOC plunges below the threshold $\underline{E}_i^{\rm th}$ to approach its lower limit $\underline{E}_i$, it is preferable to charge than discharge. If $f_i(\omega_i) > 0$, then the ESS is charged at power $P_i^e = \max\{\frac{\underline{\alpha}(E_i - \underline{E}_i)}{\underline{\alpha}(\underline{E}_i^{\rm th} - \underline{E}_i)} f_i(\omega_i), f_i(\omega_i)\} = f_i(\omega_i)$, in which case the safety filter does not intervene. In contrast, if $f_i(\omega_i) < 0$, then $P_i^e = \frac{\underline{\alpha}(E_i - \underline{E}_i)}{\underline{\alpha}(\underline{E}_i^{\rm th} - \underline{E}_i)} f_i(\omega_i)$, i.e., the ESS is discharged at a power that is smaller than $|f_i(\omega_i)|$, until the discharge power diminishes to zero as the lower SOC limit $E_i = \underline{E}_i$ is reached. A similar argument applies to the case $E_i \ge \overline{E}_i^{\rm th}$. Overall, the proposed safety filter in (7) strictly guarantees the satisfaction of the SOC limits at every time instant during the transient.

The design of $f_i(\cdot)$ can be flexible as long as it is Lipschitz continuous and its range is within $[\underline{P}_i^e,\ \overline{P}_i^e]$. In particular, $f_i(\cdot)$ can be a nonlinear mapping obtained by means of a learning method as we will elaborate. This shall make it more capable than classic linear controllers in improving transient performance, such as the maximum frequency deviation and integrated control cost.

## IV. Closed-loop System Performance

In this section, we identify conditions on the ESS controller design so that the closed-loop system (2), (6) admits desired configurations as equilibria and they are asymptotically stable.

### A. Equilibrium Set of the Closed-loop System

Let $\theta$, $\omega$, $P^m$, $P^c$, and $E$ be stacked vectors of $\theta_i$, $\omega_i$, $P_i^m$, $P_i^c$, and $E_i$ for all $i \in \mathcal{N}$, resp. Define $\delta := C^\top \theta \in \mathbb{R}^{|\mathcal{E}|}$, where the incidence matrix $C \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{E}|}$ has its $i$-th row, $k$-th column element defined as:

$$C_{ik} := \begin{cases} 1, & \text{if line } k \in \mathcal{E} \text{ starts from bus } i, \\ -1, & \text{if line } k \in \mathcal{E} \text{ ends at bus } i, \\ 0, & \text{otherwise.} \end{cases}$$

We rewrite system (2) with the feedback control (6) in the following compact form

$$\dot{\delta} = C^\top \omega, \tag{8a}$$

$$M\dot{\omega} = -D\omega - CP^b(\delta) + P^m + P^c - P^l - \mathcal{F}(\omega, E), \tag{8b}$$

$$T^m \dot{P}^m = -P^m - R^{-1}\omega, \tag{8c}$$

$$\dot{P}^c = -K^{-1}\omega, \tag{8d}$$

$$\dot{E} = \mathcal{F}(\omega, E), \tag{8e}$$

where $D$, $R$ and $K$ are resp. $\mathrm{diag}(D_1, ..., D_{|\mathcal{N}|})$, $\mathrm{diag}(R_1, ..., R_{|\mathcal{N}|})$, and $\mathrm{diag}(K_1, ..., K_{|\mathcal{N}|})$, $\mathcal{F}(\omega, E) \in \mathbb{R}^{|\mathcal{N}|}$ collects $\mathcal{F}_i(\omega_i, E_i)$ for all $i \in \mathcal{N}$, and the changes in line power flows are

$$P^b(\delta) := B\sin(\delta^0 + \delta) - B\sin(\delta^0).$$

Here $B \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{E}|}$ is the diagonal matrix collecting $B_{ij}$ for all $(i,j) \in \mathcal{E}$, and the vector-valued expression $\sin(\delta)$ takes the scalar $\sin(\cdot)$ of each element of input $\delta$. We assume the angle differences $\delta$ is within the following region[2]

$$\Delta := \left\{ \delta \in \mathbb{R}^{|\mathcal{E}|} \middle|\ \left|\delta_k^0 + \delta_k\right| < \frac{\pi}{2},\ \forall k \in \mathcal{E} \right\}. \tag{9}$$

The next result identifies the equilibrium set of (8).

**Proposition 1.** *(Equilibria of the closed-loop system): The closed-loop system* (8) *admits the following set of equilibria:*

$$\mathcal{S} = \{(\delta^*, \omega^*, P^{m*}, P^{c*}, E^*) \mid \omega^* = \mathbf{0},\ P^{m*} = \mathbf{0}, \\ P^{c*} = CP^b(\delta^*) + P^l,\ \mathcal{F}(\mathbf{0}, E^*) = \mathbf{0}\}.$$

*Proof.* It is straightforward by (8c) and (8d) that an equilibrium $(\delta^*, \omega^*, P^{m*}, P^{c*}, E^*)$ of (8) must satisfy $\omega^* = \mathbf{0}$ and $P^{m*} = \mathbf{0}$. By (8e), we have $\mathcal{F}(\omega^*, E^*) = \mathcal{F}(\mathbf{0}, E^*) = \mathbf{0}$. Then (8b) implies $P^{c*} = CP^b(\delta^*) + P^l$. □

Given our safety filter-based controller design in Section III-B, $\underline{E}_i \le E_i \le \overline{E}_i$ is strictly guaranteed. Together with (9), it follows that $\mathcal{S}$ is compact. Moreover, from (7), if $f_i(0) = 0$, then $\mathcal{F}_i(0, E_i^*) = 0$ for all $E_i^* \in \mathbb{R}$. Using this observation, we prove next the asymptotic stability of $\mathcal{S}$.

### B. Convergence to the Equilibrium Set

Let $x := (\delta, \omega, P^m, P^c, E)$. Given $x^* \in \mathcal{S}$, define the deviation $\tilde{x} := x - x^*$, and similarly for its individual elements. Following [8], we construct the energy function,

$$V(\tilde{x}) := \frac{1}{2}\tilde{\omega}^\top M\tilde{\omega} + \frac{1}{2}\left(\tilde{P}^m\right)^\top RT^m\tilde{P}^m + \frac{1}{2}\left(\tilde{P}^c\right)^\top K\tilde{P}^c \\ + \sum_{k \in \mathcal{E}} B_k \int_0^{\tilde{\delta}_k} \left[\sin(\delta_k^0 + \delta_k^* + \varrho) - \sin(\delta_k^0 + \delta_k^*)\right] d\varrho,$$

where $T^m := \mathrm{diag}(T_1^m, ..., T_{|\mathcal{N}|}^m)$ collects the generator time constants. Given (9), note that $V$ is positive definite with

[2]We allow each bus angle $\theta_i$ to change on $\mathbb{R}$, not restricting it in, e.g., $[0, 2\pi)$; however, all angle differences across lines, e.g., $(\theta_i^0 + \theta_i - \theta_j^0 - \theta_j)$, are shifted by appropriate multiples of $2\pi$ to get minimum absolute values. Furthermore, let $\mathcal{P}^l = \{P^l : \|L^\dagger\left(P^{m0} + P^{c0} - P^{l0} - P^l\right)\|_{\mathcal{E},\infty} < 1\}$, where $\|y\|_{\mathcal{E},\infty} := \max_{(i,j)\in\mathcal{E}} |y_i - y_j|$ and $L := CBC^\top$ is the Laplacian matrix. According to [19], [20], if $P^l \in \mathcal{P}^l$, then (9) holds.

respect to the origin. The time derivative of $V$ along any trajectory $\tilde{x}(t)$, $\forall t \geq 0$ of the closed-loop system (8) is:

$$\begin{aligned}\dot{V}(\tilde{x}) &= \tilde{\omega}^\top M\dot{\tilde{\omega}} + \left(\tilde{P}^m\right)^\top RT^m\dot{\tilde{P}}^m + \left(\tilde{P}^c\right)^\top K\dot{\tilde{P}}^c \\ &\quad + \sum_{k\in\mathcal{E}} B_k\left[\sin(\delta_k^0+\delta_k) - \sin(\delta_k^0+\delta_k^*)\right]\dot{\tilde{\delta}}_k \\ &= \tilde{\omega}^\top\left[\tilde{P}^m + \tilde{P}^c - \mathcal{F}(\tilde{\omega}, E) - D\tilde{\omega} - C\tilde{P}^b\right] \\ &\quad + \left(\tilde{P}^m\right)^\top R\left(-\tilde{P}^m - R^{-1}\tilde{\omega}\right) - \left(\tilde{P}^c\right)^\top\tilde{\omega} \\ &\quad + \left(\tilde{P}^b\right)^\top C^\top\tilde{\omega} \\ &= \sum_{i\in\mathcal{N}}\left[-R_i\left(\tilde{P}_i^m\right)^2 - D_i\tilde{\omega}_i^2 - \tilde{\omega}_i\mathcal{F}_i(\tilde{\omega}_i, E_i)\right], \quad (10)\end{aligned}$$

where $\tilde{P}^b := P^b(\delta) - P^b(\delta^*)$ and $\mathcal{F}(\tilde{\omega}, E) := \mathcal{F}(\omega, E) - \mathcal{F}(\omega^*, E)$. The next result establishes the asymptotic stability of the equilibrium set.

**Theorem 2.** *(Global asymptotic stability of the equilibrium set): Suppose $f_i$ in the ESS controller* (7) *is Lipschitz and non-decreasing, and $f_i(0) = 0$ for all $i \in \mathcal{N}$. Then, under* (9)*, $\mathcal{S}$ is globally asymptotically stable and the convergence of trajectories is pointwise. In particular, $\lim_{t\to+\infty}\omega(t) = \omega^* = \mathbf{0}$.*

*Proof.* From (7), whatever value $E_i$ takes, it has $\mathcal{F}_i(\omega_i^*, E_i) = 0$, and therefore $\mathcal{F}_i(\tilde{\omega}_i, E_i) = \mathcal{F}_i(\omega_i, E_i) - \mathcal{F}_i(\omega_i^*, E_i) = \mathcal{F}_i(\omega_i, E_i) = \mu_i(E_i)f_i(\omega_i) = \mu_i(E_i)f_i(\tilde{\omega}_i)$, where the state-dependent factor $\mu_i(E_i)$ belongs to $[0, 1]$ for all $i \in \mathcal{N}$. Since $f_i(\cdot)$ is non-decreasing and satisfies $f_i(0) = 0$, it follows that

$$\tilde{\omega}_i\mathcal{F}_i(\tilde{\omega}_i, E_i) = \mu_i(E_i)\tilde{\omega}_i f_i(\tilde{\omega}_i) \geq 0,$$

which guarantees that (10) is non-positive, and hence $V$ is a Lyapunov function. It implies that every equilibrium in $\mathcal{S}$ is stable. Let $\mathcal{X} := \{x \mid \omega = \mathbf{0}, P^m = \mathbf{0}\}$ be the set where $\dot{V} = 0$. By noting that $\mathcal{F}(\mathbf{0}, E) = \mathbf{0}$ no matter what value $E$ takes, cf. (7), and in combination with (8b), it is easy to see that the largest invariant set contained in $\mathcal{X}$ is $\mathcal{S}$. By LaSalle's invariance principle [21, Thm. 4.4], we conclude that $\mathcal{S}$ is globally asymptotically stable under (9) and the convergence of any trajectory of (8) is to a point in $\mathcal{S}$, cf. [22, Cor. 5.2]. □

## V. Learning to Optimize the Transient

Having identified conditions on the controller design (7) to ensure closed-loop system stability, we perform RL to find $\{f_i, \underline{E}_i^{\text{th}}, \overline{E}_i^{\text{th}}\}_{i\in\mathcal{N}}$ optimizing transient performance. Note that in principle, the $\mathcal{K}_\infty$ functions $\underline{\alpha}_i, \overline{\alpha}_i$ can also be trained by RL, see e.g. [12]. Here, we focus on optimizing $\{f_i, \underline{E}_i^{\text{th}}, \overline{E}_i^{\text{th}}\}_{i\in\mathcal{N}}$ as they have a more significant impact on transient performance. Specifically, our target optimization problem (5) becomes:

$$\min_{\{f_i, \underline{E}_i^{\text{th}}, \overline{E}_i^{\text{th}}\}_{i\in\mathcal{N}}} \sum_{i\in\mathcal{N}}\left(\|\omega_i\|_\infty + \frac{\rho}{T}\int_0^T g_i\left(P_i^e(t)\right)\mathrm{d}t\right) \quad (11a)$$

$$\text{s.t. (7), (8)}, \quad (11b)$$

$$\underline{E}_i < \underline{E}_i^{\text{th}} < \overline{E}_i^{\text{th}} < \overline{E}_i, \ \forall i \in \mathcal{N}, \quad (11c)$$

$$\underline{P}_i^e \leq f_i(\cdot) \leq \overline{P}_i^e, \ \forall i \in \mathcal{N}, \quad (11d)$$

$$\{f_i\}_{i\in\mathcal{N}} \text{ are non-decreasing}, \quad (11e)$$

$$f_i(0) = 0, \ \forall i \in \mathcal{N}. \quad (11f)$$

We use neural networks to parameterize $\{f_i, \underline{E}_i^{\text{th}}, \overline{E}_i^{\text{th}}\}_{i\in\mathcal{N}}$ such that conditions (11c)–(11f) hold. Let

$$\underline{E}_i^{\text{th}} = \underline{E}_i + s(\underline{e}_i)(\overline{E}_i - \underline{E}_i), \overline{E}_i^{\text{th}} = \underline{E}_i + s(\overline{e}_i)(\overline{E}_i - \underline{E}_i) \quad (12)$$

where $s(x) = \frac{1}{1+e^{-x}}$ is the Sigmoid function and $\underline{e}_i, \overline{e}_i \in \mathbb{R}$, $\underline{e}_i < \overline{e}_i$ are biases so that (11c) naturally holds. For $f_i$, we adopt the single-hidden-layer neural network design in [11], [12] to perform the parameterization. Specifically, let $\sigma$ represent the ReLU activation function, i.e., $\sigma(x) = \max\{0, x\}$. Each $f_i$ is parameterized using a single-hidden-layer neural network $N_i(x)$ with output projection to enforce (11d) as

$$f_i(x) = \overline{P}_i^e - \sigma(\overline{P}_i^e - N_i(x)) + \sigma(\underline{P}_i^e - N_i(x)), \quad (13a)$$

with

$$N_i(x) = \sum_{h=1}^H \overline{w}_h^i\sigma(x + \overline{b}_h^i) + \sum_{h=1}^H \underline{w}_h^i\sigma(-x + \underline{b}_h^i), \quad (13b)$$

where $\overline{b}_h^i, \underline{b}_h^i \in \mathbb{R}$ are biases and $\overline{w}_h^i, \underline{w}_h^i \in \mathbb{R}$ are weights. The next result provides the conditions on these parameters to ensure the satisfaction of (11e)–(11f).

**Proposition 3.** *(Single-hidden-layer neural network parameterization): For each $i \in \mathcal{N}$, let $f_i$ be parameterized by* (13)*. If $\overline{b}_1^i = 0$, $\overline{b}_h^i \leq \overline{b}_{h-1}^i$ (resp. $\underline{b}_1^i = 0$, $\underline{b}_h^i \leq \underline{b}_{h-1}^i$) for all $2 \leq h \leq H$, and $\sum_{h=1}^\ell \overline{w}_h^i \geq 0$ (resp. $\sum_{h=1}^\ell \underline{w}_h^i \leq 0$) for all $1 \leq \ell \leq H$, then $f_i(0) = 0$ and $f_i$ is non-decreasing. Moreover, for any non-decreasing Lipschitz function $g_i : \mathbb{R} \to \mathbb{R}$ satisfying $g_i(0) = 0$ and given any compact domain $\mathcal{J} \subset \mathbb{R}$ and $\epsilon > 0$, there exists $\overline{w}_h^i, \underline{w}_h^i, \overline{b}_h^i, \underline{b}_h^i$ and $H$ such that $|f_i(x) - g_i(x)| < \epsilon$ for all $x \in \mathcal{J}$.*

This result follows from [11, Thm. 2] and relies on the definition of ReLU function and the universal approximation property of ReLU networks. Let $\varphi := \{\overline{e}_i, \underline{e}_i, \{\overline{w}_h^i, \overline{b}_h^i, \underline{w}_h^i, \underline{b}_h^i\}_{h=1}^H\}_{i\in\mathcal{N}}$ and denote by $\mathcal{F}^\varphi$ the parameterized controller using $\varphi$. Performing a forward Euler discretization with stepsize $\Delta t$ and number of timesteps $L = \frac{T}{\Delta t}$, we can rewrite (11) in the following discrete-time parameterized form

$$\begin{aligned}\min_\varphi \ & \sum_{i\in\mathcal{N}}\left(\max_{0\leq\tau\leq L-1}|\omega_i(\tau)| + \frac{\rho}{L}\sum_{\tau=0}^{L-1} g_i\left(\mathcal{F}_i^\varphi(\omega_i(\tau), E_i(\tau))\right)\right) \\ \text{s.t. } & \delta(\tau+1) = \delta(\tau) + C^\top\omega(\tau)\Delta t \\ & \omega(\tau+1) = \left(I - DM^{-1}\Delta t\right)\omega(\tau) + M^{-1}\left(P^m(\tau) + P^c(\tau)\right. \\ & \qquad \left. - P^l - \mathcal{F}^\varphi(\omega(\tau), E(\tau)) - CP^b(\delta(\tau))\right)\Delta t \\ & P^m(\tau+1) = (I - (T^m)^{-1}\Delta t)P^m(\tau) - (RT^m)^{-1}\omega(\tau)\Delta t \\ & P^c(\tau+1) = P^c(\tau) - K^{-1}\omega(\tau)\Delta t \\ & E(\tau+1) = E(\tau) + \mathcal{F}^\varphi(\omega(\tau), E(\tau))\Delta t\end{aligned}$$

(7), (12) – (13).

We use the recurrent neural network (RNN)-based RL framework in [11] to perform the training for solving the above parametric optimization problem, where $\varphi$ are updated by gradient descent and converge to a local optimum.

**Remark 4.** *(Regularization of steady-state SOC):* In practice, it would be preferable to have the steady-state SOC, i.e., $E_i^*$, stay as close as possible to the median (zero in our case) of its feasible range $[\underline{E}_i, \overline{E}_i]$, so that the ESS is better prepared to help regulate the next disturbance. To do this, one can add a regularization term $\gamma \sum_{i\in\mathcal{N}} |E_i(L)|$ to the cost function to penalize the deviation of $E_i(L)$ from zero, at the cost of potentially degrading the performance with regards to the other terms in (5a). We illustrate the performance of this approach in the simulations below. •

## VI. Numerical Results

We conduct a case study to illustrate the performance of the proposed ESS controller. We use the IEEE 39-bus power network with 10 generators and perform Kron reduction [23] to obtain its model as (2). The inertial constants $M_i$, damping factors $D_i$, and initial power injections $P_i^{l0}$ are selected the same as [11]. We set uniform $T_i^m = 7$ seconds, $1/R_i = 10$, $1/K_i = 3$, $\overline{P}_i^e = -\underline{P}_i^e = P_i^{l0}$, and $\overline{E}_i = -\underline{E}_i = 0.3$; and we set $\mathcal{K}_\infty$ functions $\overline{\alpha}_i(x) = \underline{\alpha}_i(x) = x$ in (7), and cost functions $g_i(x) = x^2$ in (5), for all $i \in \mathcal{N}$.

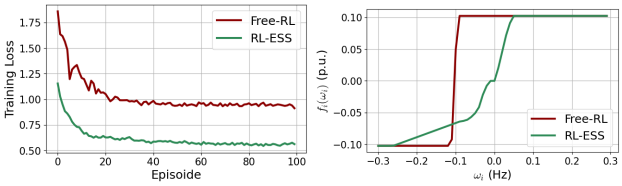
### A. Training Setup and Comparison Baseline

We build the RL environment using TensorFlow 2.7.0 and conduct the training process in Google Colab on a single TPU with 32 GB memory. We consider the time horizon of interest to be $T = 20$ seconds. To facilitate the training process, we only evaluate the first 5 seconds in training, with the discretization stepsize $\Delta t = 0.01$ seconds, i.e., $L = 500$. We set the balancing coefficient in the objective function as $\rho = 0.2$, and the number of episodes, the batch size, and the number of neurons as 100, 600, 50, resp., making the trainable parameters $\varphi$ of dimension 2020 in the learning process. We use the Adam algorithm [24] to update the parameter $\varphi$ in each episode, with a learning rate initialized as 0.1 and decaying every 20 steps on a base of 0.7. In each batch, $P_i^l$ is randomly generated in $[-1, 1]$ per unit (p.u.), and $E_i^0$ is randomly generated in $\left[0.5\underline{E}_i, 0.5\overline{E}_i\right]$ p.u., $\forall i \in \mathcal{N}$.

Similar to [11], [13], [14], we compare the proposed controller, termed RL-ESS, with the Free-RL controller. The latter is also parameterized by a single-hidden-layer ReLU neural network, using the same hyperparameters as in RL-ESS, but trained without the structural constraints (11e)-(11f) or the safety filter design (7). Instead, the Free-RL controller just disables any charging (discharging) action when the SOC is at the upper (lower) bound. This simple approach ensures satisfaction of the SOC limit, possibly with discontinuous control actions.

### B. Simulation Results

Fig. 1(a) plots the training loss curves for the proposed RL-ESS and the Free-RL methods. The loss for RL-ESS is reduced by around 46% in 100 episodes. Compared to Free-RL, the RL-ESS method has a warm start and achieves lower training loss. We also illustrate the learned control function $f_i(\omega_i)$ for generator bus $i = 4$ in Fig. 1(b).




(a) Training Loss (b) Learned control function $f_4$

Fig. 1. RL training results: (a) training loss curves; (b) learned control function $f_i(\omega_i)$ for $i = 4$.

TABLE I
Comparison of Transient Costs

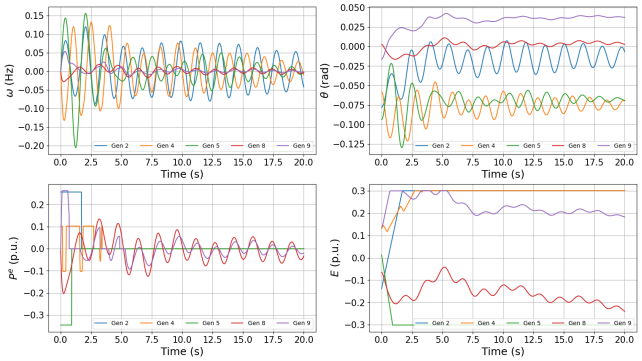
| Method | $\sum_{i\in\mathcal{N}} \|\omega_i\|_\infty$ | $\frac{\rho}{L'} \sum_{i\in\mathcal{N}} \sum_{t=0}^{L'-1} (P_i^e(t))^2$ |
|---|---|---|
| No ESS | 1.1573 | - |
| Free-RL | 0.8406 | 0.0073 |
| RL-ESS | 0.5944 | 0.0018 |

Fig. 2 compares the system dynamics for the Free-RL and RL-ESS methods under the same disturbances $P_i^l$ and initial SOCs. Compared to Free-RL, the RL-ESS method significantly enhances the transient performance with reduced frequency and power oscillations, while strictly satisfying the SOC and system stability requirements. We show in Table I the transient costs for both methods and the case without any ESS control where $L' = T/\Delta t = 2000$ (note that RL-ESS control achieves the lowest transient cost).

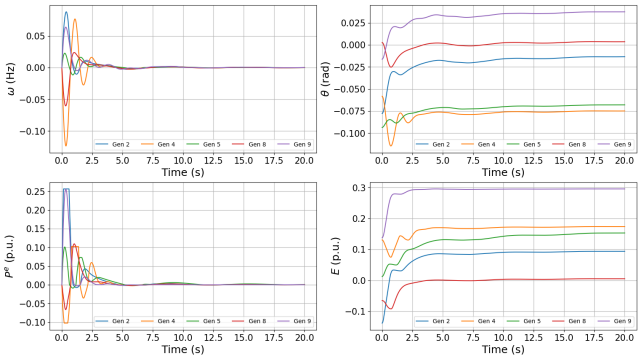
We next compare the RL-ESS with and without the regularization for steady-state SOC described in Remark 4 under the same disturbances $P_i^l$ and initial SOCs. For the regularization, we set $\gamma$ = 0.5. Fig. 3 shows the SOC dynamics for the two cases, where one can see that the regularization reduces the steady-state SOC deviation by 44.6%.

### C. Discussion

From Fig. 1(a), one can see that the RL-ESS achieves better training performance. We attribute this to the fact that RL-ESS optimizes the control policy over a reduced controller class leveraging model structure information, which helps to find a better local optimum under the same training setting. We also observe in Fig. 2 that Free-RL leads to large oscillations of the system state. The reason is twofold. First, the Free-RL control lacks the safety filter design and thus makes the control action suddenly jump when the SOC reaches its boundary, causing the oscillations. Second, without enforcing $f_i(0) = 0$, cf. Fig. 1(b), the Free-RL control does not guarantee the control action to vanish at the equilibrium, leading to possible oscillations around the equilibrium. This observation highlights the importance of our safety filter design and stability condition for the proposed ESS controller.

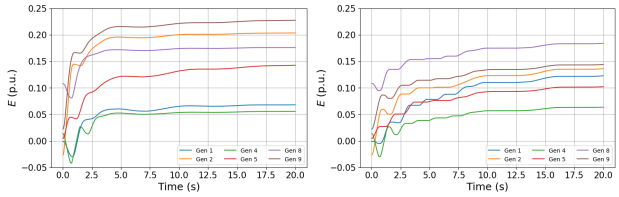



(a) Free-RL control




(b) RL-ESS control

Fig. 2. System dynamics under (a) Free-RL and (b) RL-ESS controls.




(a) Without regularization (b) With regularization

Fig. 3. SOC dynamics for RL-ESS control (a) without and (b) with regularization. The value of the steady-state SOC deviation $\gamma \sum_{i \in \mathcal{N}} |E_i(L')|$ is resp. (a) 0.6522 and (b) 0.3616.

## VII. Conclusions

We have proposed a decentralized controller for energy storage systems (ESS) to help stabilize frequency deviations under disturbances. The proposed controller contains a novel design of a safety filter to enforce the ESSs' state of charge constraint during the transient. We have identified conditions on the controller design to guarantee that the closed-loop system converges to a desired equilibrium with nominal frequency. These conditions are then imposed in neural-network training to synthesize decentralized controllers through reinforcement learning. Simulations show that the RL-based ESS controller effectively reduces frequency oscillations and transient control cost. Future work will consider more realistic power network scenarions with pure load buses, compare the proposed design to other ESS control strategies, including standard ones, explore other forms of controller design, study their corresponding stability conditions, and leverage distributed communication among controllers.

## References


[1] Y. Jiang, E. Cohn, P. Vorobev, and E. Mallada, "Storage-based frequency shaping control," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5006–5019, 2021.

[2] A. Jokić, M. Lazar, and P. P. J. van den Bosch, "Real-time control of power systems using nodal prices," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 9, pp. 522–530, 2009.

[3] A. Jokić, M. Lazar, and P. P. J. van den Bosch, "On constrained steady-state regulation: Dynamic KKT controllers," *IEEE Transactions on Automatic Control*, vol. 54, no. 9, pp. 2250–2254, 2009.

[4] N. Li, C. Zhao, and L. Chen, "Connecting automatic generation control and economic dispatch from an optimization view," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 3, pp. 254–264, 2016.

[5] F. Liu, Z. Wang, C. Zhao, and P. Yang, *Merging Optimization and Control in Power Systems: Physical and Cyber Restrictions in Distributed Frequency Control and Beyond*. John Wiley & Sons, 2022.

[6] C. Zhao, U. Topcu, N. Li, and S. H. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1177–1189, 2014.

[7] E. Mallada, C. Zhao, and S. H. Low, "Optimal load-side control for frequency regulation in smart grids," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6294–6309, 2017.

[8] C. Zhao, E. Mallada, S. H. Low, and J. Bialek, "Distributed plug-and-play optimal generator and load control for power system frequency regulation," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 1–12, 2018.

[9] Z. Wang, F. Liu, S. H. Low, P. Yang, and S. Mei, "Distributed load-side control: Coping with variation of renewable generations," *Automatica*, vol. 109, no. 108556, 2019.

[10] Z. Wang, F. Liu, C. Zhao, Z. Ma, and W. Wei, "Distributed optimal load frequency control considering nonsmooth cost functions," *Systems & Control Letters*, vol. 136, no. 104607, 2020.

[11] W. Cui, Y. Jiang, and B. Zhang, "Reinforcement learning for optimal primary frequency control: A Lyapunov approach," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1676–1688, 2023.

[12] Z. Yuan, C. Zhao, and J. Cortés, "Reinforcement learning for distributed transient frequency control with stability and safety guarantees," *Systems & Control Letters*, 2022, submitted.

[13] Y. Jiang, W. Cui, B. Zhang, and J. Cortés, "Stable reinforcement learning for optimal frequency control: A distributed averaging-based integral approach," *IEEE Open Journal of Control Systems*, vol. 1, pp. 194–209, 2022.

[14] W. Cui, G. Shi, Y. Shi, and B. Zhang, "Leveraging predictions in power system frequency control: an adaptive approach," in *IEEE Conference on Decision and Control*, Marina Bay Sands, Singapore, 2023.

[15] A. R. Bergen and V. Vijay, *Power System Analysis*. Prentice-Hall, 2000.

[16] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Reinforcement learning for selective key applications in power systems: Recent advances and future challenges," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2935–2958, 2022.

[17] F. S. Gorostiza and F. M. Gonzalez-Longatt, "Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5039–5050, 2020.

[18] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1653–1656, 2018.

[19] Y. Zhang and J. Cortés, "Distributed transient frequency control for power networks with stability and performance guarantees," *Automatica*, vol. 105, pp. 274–285, 2019.

[20] F. Dörfler, M. Chertkov, and F. Bullo, "Synchronization in complex oscillator networks and smart grids," *Proceedings of the National Academy of Sciences*, vol. 110, no. 6, pp. 2005–2010, 2013.

[21] H. Khalil, *Nonlinear Systems, 3rd ed.* Prentice Hall, 2002.

[22] S. P. Bhat and D. S. Bernstein, "Nontangency-based Lyapunov tests for convergence and stability in systems having a continuum of equilibria," *SIAM Journal on Control and Optimization*, vol. 42, no. 5, pp. 1745–1775, 2003.

[23] F. Dörfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 1, pp. 150–163, 2012.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.