We propose a method to minimize the long-term cost of energy generation while improving grid stability. Currently, the cost of energy generation is minimized myopically (day by day) via the economic dispatch problem, which i) does not internalize the effects of generation variability, ii) does not account for the long-term effects of losing too many existing (paid off) conventional plants, and iii) has the detrimental impact of not systematically maintaining grid inertia. The current dispatch solution favors low cost but inherently more variable renewables, which require intermittent back-up from either conventional or expensive peakers.

We first propose our Augmented Dispatch for Inertia method which incorporates the cost of maintaining grid inertia stability directly in the economic dispatch selection, thus more accurately capturing the impact of renewable energy growth and conventional plant retirements. Second, to address the long-term loss of conventional plants due to their underuse, we propose our Balanced Dispatch algorithm that selects key, future-needed conventional generators with enough frequency to maintain their viability. We show via simulation that our methods result in substantially lower long-term generation cost and a notable increase in grid resilience.

1 INTRODUCTION

Power generation is currently managed by Independent System Operators (ISO) in a myopic way (day by day). The goal of ISOs is to clear the market so that the load is met using the cheapest available generation while maintaining grid stability. Thus, less expensive sources of energy like Wind are typically being selected first. In the longer term, this myopic selection leads to the slow loss of conventional generation, due to under-use, as more and more renewables penetrate the system [6]. Additionally, in the short-term, when the wind forecast is not accurate, system operators have to either activate reserves or use quick ramping peaker plants.

In both of these cases, conventional sources like Coal and Gas, that are moderately priced and have longer ramp-up times, are being selected less often. This reduces their economic viability, slowly driving them out of business. This reduced use and subsequent loss of many conventional plants pushes the system to the two extremes: cheap, variable Wind and expensive, quick ramping Peakers, resulting in i) more extreme price fluctuation in the system, ii) lower overall system inertia stability (without continuous monitoring and action by the ISO), iii) an increased overall long-term generation costs. It also results in more frequent cases of scarcity pricing [5, 16].

Moreover, regarding overall system resilience, conventional generation provides rotation mass in the grid, called \textit{inertia}, which acts as a buffer to slow down the rate of frequency change. As seen in a number of events in recent years, when there is a dramatic penetration of renewable generation into a system without adequate inertial rotating mass, the system can be susceptible to poor power quality and blackouts [24]. Conventional usage, unlike renewables’ usage, bolsters electrical Inertia to help maintain a stable and reliable grid. Additionally, a higher availability of conventional capacity in the long-term reduces the required need for peaker plants when renewable generation is low\(^1\), which keeps overall generation cost lower.

The goal of this paper is to study the implications of the current economic dispatch methods on both the Long-Term Cost of Generation and on Grid Inertial Stability, and to offer potential solutions. Specifically, we propose two complimentary approaches, with the first (i) yielding a myopic, short-term solution to systematically ensure satisfactory grid Inertia, and the second (ii) presenting a balanced dispatch solution to stabilize long-term generation cost via a minimum threshold of non-variable generation. For the former, we address the issue of the continuous violation of grid Inertia (starting in 2022) under the current economic dispatch

\(^1\) These periods of low renewable generation are in reference to day over day operation periods and not inter-day forecast errors periods.
method ("Economic Dispatch"), and offer a solution to ensure grid inertia is always maintained above the minimum Inertia threshold. For the latter, we propose a threshold algorithm that selects key conventional sources often enough to maintain their viability, even when they are not among the cheapest generators within that period. Ultimately, our goal is to offer a systematic approach for finding the most cost-effective way to produce power in the long term while still maintaining minimum Inertia and other grid reliability constraints.

We give the following historic example to illustrate how the shortage of conventional generation can have devastating consequences. In 2011 two Coal plants in ERCOT shut down for two days because of a snow storm and the remaining generation could not meet the load, resulting in rolling blackouts and a sudden rise in energy price from around $30/MWh to $3,000/MWh (the maximum allowable at that time by the Texas ISO, The Electric Reliability Council of Texas, "ERCOT"). Additionally, in the summer of 2011, a record heat wave pushed the ERCOT load to an all-time high, at which point there was again no more generation available to meet the load. Market prices peaked at $3,000/MWh for only 28 of 8760 hours during these two periods, but drove the average price that year up by 30% (from $38/MWh to $50/MWh). These historic episodes give a glimpse into the future when too many conventional plants have shut down due to under-use and are not available during times of need.

To address the shortages in 2011, ERCOT raised the market price cap from $3,000 to $9,000/MWh. With this move not being enough to prevent losses of capacity reserves [6, 21], in 2014, ERCOT additionally implemented a mechanism called scarcity pricing [5, 16]. These together are proving moderately effective, but i) are disproportionately promoting high-priced peakers and low-priced renewables to build new generation over conventional [19] and ii) they only start having noticeable effects a few years before larger issues tend to occur, which does not allow non-peaker conventional plants enough lead time to justify entering the market. This polarizes the system and leads it to higher long-term generation costs and lower grid inertia stability.

ERCOT, and other system operators, thinks pure economic incentive and free market forces will drive an optimal generation mix to the best long-term cost. However, we demonstrate in this paper that this is not necessarily the case. Consider a simple example where the system operator is offered the following sources in Fig. 1: an inexpensive variable source at $25/MWh with capacity 250MW (green/wind), a moderately priced conventional source at $35/MWh with capacity 100MW (blue/conventional), and an expensive peaker source at $150/MWh (red/peaker) with unlimited capacity.

Assume that the conventional plant has not been picked by the system operator for a while and it will be replaced by a 100MW renewable source if not picked in the current round. If the demand is 200MW in both rounds, economic dispatch will not assign any generation to the conventional plant in round one (Fig. 1-(top)). Now, in round two, when renewable generation is lower, the two variable sources only produce 100MW and 50MW. Therefore, in the absence of a conventional source, the system operator has to schedule the remaining 50MW with the peaker source.

On the other hand, if we are aware of the conventional’s situation, we will select it before the variable source (Fig. 1-(bottom)) in round one. This will increase the generation cost in round one, but the survival of the conventional will help the system operator avoid the use of the peaker in round two. The average generation cost is $40.6 in the first case (myopic economic dispatch), and $30 in the second case (considering longer-term effects). To conclude, we are able to reduce the long-term generation cost relative to the standard economic dispatch solution, by trading off slightly higher cost now for lower cost in the future.

Therefore, we raise the questions: What if we took a proactive approach in maintaining the existing needed generation sources in a system? What if instead of having the market price rise to keep future-needed conventional generators in business, we choose to promote their viability in a different way? The system could then operate with a more balanced mix of cheap renewables, mid-priced conventional and high-priced peakers, while providing a lower long-term generation cost and better stability.

**Summary of contributions:** First, we propose an Augmented Dispatch algorithm that incorporates the minimum inertia requirement into the economic dispatch solution. Additionally, we propose a separate Balanced Dispatch algorithm that keeps a certain level of conventional generation in the system by dispatching key conventional plants in a way to maintain their viability. Third, we consider the combination of the two algorithms (Augmented Balanced Dispatch). All three approaches achieve a lower long-term cost, while improving the grid inertia but the combined approach achieves the best results. We examine the effectiveness of our proposed algorithms via simulation over the time horizon of 2019-2055 using the available data for the ERCOT grid.

### 2 RELATED WORK

Our study of the change in generation mix over time is related to prior work on the effect of Renewable integration [15]. Green and Vasilakos [3] show that large amounts of intermittent Renewables in Great Britain would shift the generation mix towards power stations with lower overhead costs. Traber [25] also studies the effects of differing capacity policies (such as capacity markets or reserve obligations) in central European electricity markets, and shows that such changes depend on the existing power plant mix as well as the elasticity of the demand.
Several approaches have been proposed [2, 14] to calculate the cost of wind integration. Hirth et al. [4] define the integration cost as the system-level cost of wind and solar generators from their temporal variability, uncertainty, and location constraints; and show that at a renewable penetration of 30–40%, the integration cost can be up to 50% of the generation cost. To address these costs and support the large-scale integration of renewable energy, several solutions have been proposed for optimal structuring of reserves [10, 12, 30].

In addition, Lorca and Sun [11] and Wei et al. [26] have proposed robust optimization techniques for cases of high wind penetration to tackle the huge reliability concerns.

Another concern regarding the increased penetration of renewable generation is the reduced inertia of the system [29]. Several methods have been proposed to imitate the kinetic inertia of conventional [1, 31]; however, current wind and solar technologies cannot compete with conventional generation in offering certain ancillary services, specifically those related to primary frequency control [17].

In this paper, we aim to address the aforementioned optimal generation mix and inertia violations concerns within the economic dispatch problem. Even though the long-term consequences of wind integration have been studied in the literature [22, 27, 28], our work differs from prior literature as it takes a proactive approach to solve the problem by integrating the costs of maintaining grid inertia and conventional plant retirement issues into economic dispatch.

3 MODEL AND PRELIMINARIES

We propose a new process for selecting energy generation, which combines traditional economic dispatch (described next) with an additional threshold algorithm that maintains a minimal amount of conventional generation (see Section 5).

The traditional economic dispatch problem ("Economic Dispatch") aims to minimize the cost of generation subject to transmission and operational constraints [8]. Each day generators provide the ISO their maximum capacity and corresponding energy cost function. Allocating energy generation is then done by solving Economic Dispatch, and the Locational Marginal Prices (LMP) are derived as part of the Economic Dispatch solution.

Let \( x_i \) denote the power generated by source \( i \), and \( f_i(x_i) \) be the associated cost. Then the goal is to solve the following optimization problem:

\[
\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} f_i(x_i) \right\} \text{s.t.} \sum_{i=1}^{n} P_i \leq x \leq \bar{P}, \ g(x) = 0, \ h(x) \leq \bar{h}, \]

where \( n \) is the number of generators, \( x \) is the vector of all generation levels \( (x_i)'s) \), \( P \) and \( \bar{P} \) are the vectors of lower and upper generation limits, \( g(x) = 0 \) captures all equality constraints (e.g., meeting the demand), and \( h(x) \leq \bar{h} \) models the network inequality constraints (e.g., thermal flow limits). Assuming a linear generation cost \( f_i(x_i) = c_i x_i \), and employing the DC power flow model, optimization problem (1) simplifies to a linear program (LP).

In the absence of network congestion, the optimization problem (1) reduces to the greedy assignment of power generation from the least expensive source upwards until the demand is met, as demonstrated in the following example.

**Example 3.1.** Consider a four-bus network as shown in Fig. 2, where all the lines have the same impedances. Assume there is a load of \( D_0 = 1000\text{MW} \) at bus 0, and three generators of low, medium, and high price at buses 1, 2, and 3, respectively. In particular, we have the following generation costs and capacities:

- \( c_1 = $25/\text{MWh}, \ P_1 \in [0, 800] \)
- \( c_2 = $35/\text{MWh}, \ P_2 \in [0, 500] \)
- \( c_3 = $150/\text{MWh}, \ P_3 \in [0, 1000] \)

Now assume that the line between buses 2 and 3 has a capacity of 250MW (Fig. 2-(left)). Writing the DC power flow equations, this requires that \( \frac{P_1}{P_2} - \frac{P_2}{P_3} \leq 250\text{MW} \). Minimizing \( c_1 P_1 + c_2 P_2 + c_3 P_3 \) subject to this constraint as well as power balance constraints and generation limits, we get \( P_1^* = 700, P_2^* = 200, \) and \( P_3^* = 100\text{MW} \). Further, the dual variable corresponding to the power balance constraint at each bus determines the LMP at that bus. The LMP at each bus can also be interpreted as the marginal cost of an extra unit of energy, if we add an infinitesimal load at that bus.

On the other hand, if we do not have any congestion in the network (e.g., when the line limits are higher than the actual power flows, as in Fig. 2-(right)), solving the above optimization problem reduces to assigning full generation to sources in the order of increasing cost, until the demand is met. In this example, \( P_1 \) will generate at full capacity (800MW) and the remaining 200MW will be provided by \( P_2 \). In the absence of congestion, the LMP will be the same at all nodes and equal to the price of the last assigned source \( (c_2 = $35/\text{MWh} \) in this case). The greedy assignment of this Example 3.1 is formally stated in the following proposition.

**Proposition 3.2.** The Economic Dispatch solution in the absence of network constraints sorts all energy sources in increasing order of generation costs \( c_i \). Then, it finds the smallest index \( k \) such that the sum of the source capacities \( \sum_{i=1}^{k} P_i \) is greater than or equal to the demand.

In the rest of this paper, we assume a network model with no congestion; however, we do expect our findings to hold for the more general case with possible congestion. For example, in many systems, the good portion of congestion is due to the growth of renewables near the edges of the system (e.g., far West Texas), and both of our algorithms in this paper prioritize a subset of conventional generation over renewables. Since conventional generation is typically more centralized within renewables than renewables, finding a
subset that maintains or even improves congestion is not difficult in most cases. Therefore, one can solve Economic Dispatch with or without congestion after our algorithm(s) assigns its necessary generation to underused conventials.

We classify the sources of energy into three groups:
- **variable sources** – renewables, e.g., Wind or Solar;
- **conventional sources** – fossil fuel energy sources, e.g., Nuclear, Coal, Gas;
- **peaker sources** – natural Gas plants that choose to largely operate during higher demand periods and/or that are capable of rapidly ramping.

For the remainder of this paper we refer to conventional sources as "Convs". Variable sources are typically connected to the grid through converters, making their operation independent of system frequency, which results in reduced system inertia [29]. Therefore, the increased penetration of renewables brings a new challenge to grid inertia stability, discussed in the next section.

4 INERTIA AND GRID RESILIENCE

4.1 Inertia background

One consequence of high renewable penetration and thus periodic low Convs usage is low system Inertia. Electrical inertia helps oppose the changes in current and acts as a buffer against rapid frequency change. In order for a generator to contribute to system Inertia, it has to have a direct electromagnetic coupling between the power system and itself, which allows system disturbances to be translated into mechanical torque that acts on the generator’s rotor [23]. If frequency is not maintained within its operating band, part of the network can be forced to shut down. This causes severe strain on the rest of the system. If not dealt with effectively, cascading outages can quickly lead to a major blackout such as that experienced in the US Northeast in 2003, when 60 million people ended up without power [24]. The 2019 major blackouts in New York City and the U.K. further emphasize the need for greater grid resilience [24].

It is well-known in the power systems field that there is an important connection between system Inertia and grid stability. However, as recognized by Johnson et al. [7], few studies have explored the future stability issues that may result from renewables growth and power plant retirements. ERCOT presents the correlation between Wind penetration and system Inertia in the ERCOT grid, as seen in Fig. 3 [18]. In the figure, the data points progress to the right each year, indicating the increasing renewable penetration. Meanwhile, the system Inertia slowly drops (y-axis) towards the minimum Inertia threshold. This results largely from renewables’ growth in recent years notably outpacing demand growth [18].

Based on the above evidence, we now investigate how the increasing growth of non-Inertia adding renewables and the economically-forced retirements of Convs impact Inertia. Per a 2019 ERCOT study, ERCOT’s minimum Inertia for its system to function safely is 94 GWs. Fig. 4 explains the computation for this critical Inertia value [13]. The basics behind this are: ERCOT assumes a worst case scenario that its two largest plants (two 1280 MW Nuclear plants) could trip offline at exactly the same moment and that its system has to be able to maintain frequency about 59.7 hertz for 0.42 second to allow time for emergency load shedding actions to take effect. A major way that most ISOs accomplish frequency stability is through the use of rotational Inertia. When manufactured, every generator has an Inertia constant $x$, meaning that the usage of 1 MW of this generator produces $x$ MWs of Inertia. We refer the reader to Appendix A for the Inertia constants of different types of power plants. Since we do not have proprietary access to ERCOT individual plants Inertia constants and since most Convs in our simulation are Gas combined-cycle (see Appendix B), we assume an Inertia constant of 5 for all Convs, which is both the weighted average of all the Convs and the midpoint of the combined-cycle 1.1 to 9 Inertia constant range detailed in Appendix A. With detailed plant Inertia constant values, our model is able to exactly output the best Convs to use each day.

Next, we examine via simulation the severe Inertia violations that are expected to arise in the ERCOT system under the current Economic Dispatch method, and then introduce our ADi Algorithm (hereafter defined), a method that enables healthy Inertia grid resilience.
4.2 Inertia simulation

Fig. 5 shows the progression of plant usage under Economic Dispatch, along with average annual Inertia, the worst 10 days Inertia average and the 90th worst Inertia day each year. ERCOT’s lowest Inertia days are above the minimum threshold in Year 2019. However, based on ERCOT data [21], there are around 14 GWs of Wind and 7 GWs of Solar planned to be built in years 2019-2021, which is expected to result in the sudden drop of Inertia starting in year 2022. Both the 90 worst days and the worst 10 days of Inertia under current Economic Dispatch are expected to remain below the 94 GWs threshold from 2026 onward, which strongly suggests that the ERCOT system under its current methodology does not have enough Inertia resilience to counteract this rapid insertion of 20+ GWs of new renewables.

An effective myopic method to prevent falling under the Inertia threshold is to enforce enough Convs usage which we term Augmented Usage. Since the current method does not have an established systematic approach to plan ahead for this needed use of Convs, and because not every Conv plant desires to be operationally committed in the Day-Ahead market versus just financially committed, we propose integrating compensation into selected plants’ payments for their operational commitment of Inertia support. We thus designate the corresponding unit cost of selecting the needed Convs as 50% higher than if they were not selected (i.e. not additionally committed for Inertia support). Example: A plant is chosen at $36 per MWh, but to additionally designate and use it for Inertia support, it is paid $54 ($36*1.5) per MWh instead. Further, for days we have violations, we increase the Inertia requirement to 110 GWs from the 94 GWs threshold, effectively internalizing a small buffer.

We call the the algorithm described above Augmented Dispatch for Inertia (ADI) and present it as Algorithm 1.

Algorithm 1: Augmented Dispatch for Inertia (ADI)

| Input: a dispatching algorithm DispatchAlg. |
| Initialize capacity factors & demand for each day; |
| for each day in the time horizon T do |
| Run DispatchAlg; |
| Compute Inertia; |
| If Inertia < 94 GWs then |
| Clear the previous plant selection; |
| While Inertia < 110 GWs do |
| Randomly select Convs and give them 50% higher prices; |
| End |
| While demand is not met do |
| Run DispatchAlg with remaining plants; |
| End |
| End |

Fig. 6 shows the usage progression after applying the ADi Algorithm with the violet bars representing the Augmented Usage. As one can see, the worst 10 day ave line is now always kept above the Inertia threshold. Effectively, a healthy grid Inertia resilience is maintained by applying the ADi algorithm to the current Economic Dispatch (“Augmented Economic Dispatch”).

Moreover, Fig. 7 shows the comparison between the average annual cost with violations (blue curve, Economic Dispatch) and the average annual cost with violations fixed (purple curve, Augmented Economic Dispatch). The average cost with Augmented Economic Dispatch is slightly higher than Economic Dispatch for the first 15 years, but stays lower after Year 2034. This is largely because with the forced Augmented Usage, the system is able to save more initial Convs in the early years, which in turn reduces the need for some higher priced Peaker and new higher cost Convs usage in the early years.
later years on days when the variable sources are unable to provide enough power. This statement is further reinforced in Fig. 8, where our Augmented Economic Dispatch (in the process of maintaining minimum Inertia) ends up saving about 15 more initial Convs than current Economic Dispatch does. These saved initial Convs are the key reason that the ADi Algorithm is able to save an average of $5 per MWh (or 11%) in the later years of the model.

Now, consider an alternate approach. If we were able to plan ahead by maintaining the viability of a selected number of initial Convs, we could hugely decrease or even avoid these extra costs associated with Inertia +50% provision incentives discussed above. In the next section, we propose a long-term economic dispatch method that we call Balanced Dispatch (detailed hereafter). By using a more balanced generation mix, our Balanced Dispatch algorithm substantially solves the above Inertia violation problem at very minimal cost while concurrently producing a much lower long-term average generation cost (see Section 7).

5 BALANCED ECONOMIC DISPATCH

We now turn to the long-term implications of the current Economic Dispatch, with a focus on the economic viability of conventional generation. Under the current Economic Dispatch, low-cost renewable generation is favored over moderately priced Convs. Given that Convs can have significant overhead in fixed maintenance requirements, full-time employees, capital cost repayment, etc. [9], their cost of operation only partially decreases if they are not being utilized; while their revenues decrease linearly. This results in Convs closing down [5, 16], which in turn increases the need for and use of Peakers, ultimately increasing the cost of generation in the long-run as shown in our simulations in Section 6.

We therefore introduce a more balanced dispatching algorithm called Balanced Dispatch, which models the process as follows. We start with a fixed number of generators \( n \), initialized to three general types: (i) Renewables, which are cheaper but require some level of more expensive backup, (ii) Convs, which are moderately priced, and have a steadier output and iii) Peakers, which are high priced.

The cost of each plant’s energy production is chosen at the start of the simulation from a uniform distribution based on the projected real-world prices for each corresponding generator type [9]. While the capacity of the Convs is assumed fixed, the capacity of the renewable sources varies from day to day depending on the forecast and season. Similarly, the demand \( d' \) varies from day to day and is independently drawn from a truncated normal distribution centered around the average demand forecast. We assume a finite horizon of \( T \) time periods and are interested in minimizing the sum of generation costs across all periods.

Among the various power plants, there are specific types that are designated as base-supply, which means they are selected and used every day. In our model, Solar and Nuclear power plants are base-supply. This is because Solar is typically distributed, utility in-house or under contract, and Nuclear is traditionally used as base-supply.

Furthermore, we model and monitor the economic viability of each conventional. We do this by the system operator considering a Window \( W \) of consecutive periods: If the plant fails to be selected for \( \gamma W \) of the periods in any Window \( W \) (where \( \gamma \leq 1 \) is an activity requirement based on each plant’s net revenue needs for the periods) then it is put on a three-year “probation” with the probability to go out of business in each year of the probation being \( 1/3, 2/3, \) and 1 respectively (simulating a time-frame for the generator to make a shutdown timing decision). If during the probation, the plant is selected for more than \( \gamma W \) periods of the Window \( W \), then its probation is cleared. Ultimately, whenever a plant goes out of business, it is removed and gets replaced by a combination of new variable and some conventional sources (similarly to current market practices).

Since we have distributional access to future forecasts, we could calculate the optimal solution using reverse dynamic programming. This solution though would be computationally expensive and we are interested in a simpler threshold algorithm instead that keeps a needed amount of Convs.

In our Balanced Dispatch method, we maintain the total capacity of Convs above a given percentage threshold (denoted \( \beta \)) of the average demand. We do this annually by flagging Convs from lowest cost up until the \( \beta \) threshold is reached. Then, each day, we dispatch the flagged Convs that are close to their viability requirement (\( \gamma \)) before dispatching any other sources. We explain how \( \beta \) is chosen in Section 6. This algorithm is summarized as Algorithm 2.
We construct our simulation based on real-world existing and planned plant data provided by ERCOT [21]. Our simulation is run every day from year 2019 through year 2055. We divide the existing and planned generators (with their respective capacities) into specific sub-types: Wind, coastal Wind, Solar, Nuclear, Coal, Gas and Gas-Peakers. In the year of 2019, there are $n = 287$ generators in ERCOT, including 96 ‘inland’ Wind, 12 coastal Wind, 38 Solar, 4 Nuclear, 22 Coal, 64 Gas (mostly larger combined-cycles), and 51 Gas Peakers (mostly non-combined-cycles). We refer the reader to Appendix B for the 2019 list of installed capacities of the different generator types in ERCOT.

In Table 1, we provide the price ranges we use for each of these plant types and we further subdivide them by age to account for the extra income needed to support both capital investment costs early in a plant’s life and for increased maintenance, operation and efficiency costs in older plants. The prices of the plants are derived from a combination of the US levelized cost of energy calculated and projected by Lazard [9] and one of the author’s inside industry knowledge of ERCOT pricing (see Appendix C and Fig. 14 for more details on the price setup). We derived the price ranges in Table 1, but our model can also utilize any provided price ranges.

At the start of each simulation, each plant’s unit price is chosen from a uniform distribution with bounds specified in the ‘Price’ column of Table 1. We also add an infinity Peaker with unit cost $195 for when there are no other remaining plants available, thus mimicking the real-world when systems are near their limits [6, 21]. In that near-limits situation, the average market price would move towards the price cap (e.g. $9,000 MWh in ERCOT) and mechanisms like scarcity pricing [5, 16] would kick in more often, thus enticing new plants such as fast- and easy-to-construct Peakers to be built.

For variable sources, the capacity factors determine their actual daily capacity. In our model, there are three random capacity factors: ‘inland’ Wind factor, Coastal Wind factor, and Solar factor. Wind factor and Coastal Wind factor follow empirical distributions modelled after a real-world north-ERCOT wind farm (see Appendix D for details on capacity factor distributions). Additionally, we utilize a Wind seasonal factor also based on the same North Texas wind farm to mimic real-world weather patterns, with 1.333 for spring, 0.667 for summer, and 1.0 for the fall and winter. Our Solar factor follows a truncated normal distribution centered at 0.5 with a seasonal factor of 1.0 for spring and fall, 1.333 for summer, and 0.667 for winter. These random capacity and seasonal factors scale Wind and Solar productions daily, capturing in this way the variability of Renewables.

For demand values, ERCOT provides the average and peak monthly demand forecasts for years 2019 through 2028. In our model, the demand for each day is drawn according to a truncated normal distribution that is centered at the corresponding ERCOT-provided average monthly demand for that year [20]. For details of the demand setup, we refer the reader to Appendix D.

Finally, at the beginning of each year, our model adds new plants to support demand increase. The demand growth is defined as the maximum monthly demand increase between the current year and the previous year. Our model builds three types of new plants to share the growth demand: Gas, Wind and Solar, with Gas plants built with a total capacity of 70% of demand growth, Wind built with a total capacity of 60% and new Solar with a total capacity of 10%. These together equate to ~100% new plant coverage of demand growth ($0.7 + 0.6 \times 0.38 + 0.1 \times 0.5$). The average 0.38 Wind and 0.5 Solar capacity factors are detailed in Appendix D.
6.2 Simulation results

Fig. 9 shows the evolution of the initial Convs over time, for the Economic Dispatch as well as our Balanced Dispatch with different threshold percentages $\beta$. We set the viability Window $W = 3$ years, and also $y = 0.5$, meaning that if a Conv generator is not chosen by the ISO in at least half of the days in a 3-year Window, it will be put on a three-year-long "probation" to determine if it will retire permanently or not (as described in Section 5). When a Conv plant retires, it is replaced with three new generators: Wind with 60% of the retired plant’s capacity, Solar with 10% of retired capacity, and Gas with 30% of retired capacity. As one can see in Fig. 9, only around one-third of the initial 97 Convs survive under the current Economic Dispatch method. In contrast, Balanced Dispatch stops the loss of Convs when the total Conventional capacity drops to the chosen threshold $\beta$, by dispatching the Convs whose viability is threatened, even if they are not the cheapest dispatchable sources at that time.

Fig. 10 shows for average generation cost, in the short-term, Balanced Dispatch matches and in the long-term, Balanced Dispatch strongly outperforms current Economic Dispatch. We have plotted the annual averages to get smoother curves for better visual comparison. In Fig. 5, we see that the system uses an increasing amount of Peakers (the top red bars) to meet the daily demands, which is the major reason for the current Economic Dispatch’s notably higher average generation cost (blue curve) in Fig. 10. Meanwhile, if we save more Convs than we need (purple and red curves), the cost also goes up, but it is still cheaper than when we do not save enough Convs (orange and blue curves). The optimal threshold ($\beta$) can be found via binary search, as shown in Appendix E, and is around 85% in our modelled case (i.e., the green curve in Fig. 10). As indicated in Fig. 11, $\beta = 85\%$ leads to a healthy mix of variable and conventional sources, which results in both a lower average generation cost and a more predictable supply of power.

We understand the concern that with Balanced Dispatch artificially selecting a number of Convs to use each year, the competition in energy generation may become more restricted. Our method though minimally interferes with the overall dispatch selection by saving only a small percentage of Convs each year. We define “saved” Convs as the plants: i) that would have retired under the current Economic Dispatch method and/or ii) that are selected for use by Balanced Dispatch prior to other sources. It is important to note that under Balanced Dispatch, we do not fix any plants to be permanently saved. When saved, a plant is only saved for that single year and is available to retire the next year.

It turns out in Balanced Dispatch at $\beta = 85\%$, we only artificially select $\sim 12$ initial Convs each year. This shows that a small artificial selection of 12 (of 287) plants can have a huge impact on system performance in terms of both cost and embedded Inertia resilience (as detailed in Fig. 10 & Table 2).

7 BALANCED DISPATCH WITH INERTIA CONSIDERATION

In this section, we examine the system Inertia under Balanced Dispatch. We show that by using Balanced Dispatch, the severe Inertia problem present under the current Economic Dispatch (in Fig. 5) dramatically improves. Fig. 11 shows the plant usage progression for Balanced Dispatch with the optimal $\beta = 85\%$. The black dashed line (average annual 10 worst days of Inertia) is only below the minimum threshold between years 2022 and 2030, and is always above the threshold after year 2035. This is in firm contrast to current Economic Dispatch that has major Inertia threshold violations throughout the entire model in all years between 2022 and 2055 as shown in Fig. 5. Alternately after year 2035, under Balanced Dispatch, the system is resilient enough to continuously maintain required Inertia without any additional inertial support (as described in Section 4). In Table 2, we count the total number of days where the Inertia falls under the 94 GWs required minimum threshold (from Section 4), and divide it by the number of simulation years. We can see that on average, the current Economic Dispatch has a grid Inertia below the threshold for almost 1/3 of each year (114 out of 365 days). Larger $\beta$’s have zero days below the threshold and Balanced Dispatch at the optimal $\beta = 85\%$ only has an average of 7 violation days per year.

Figure 9: Long-term loss of Convs: Current Economic Dispatch versus Balanced Dispatch at different $\beta$ settings.

Figure 10: Average generation cost: Current Economic Dispatch versus Balanced Dispatch at different values of $\beta$. 

We assume the price ranges for generators in Table 1 already take $y$ values into account, meaning that the LCOE values (detailed in Appendix C) already internalize plant usage frequencies.
Figure 11: Detailed plant usage at $\beta = 85\%$ with proposed Balanced Dispatch and worst 10 days ave Inertia curve.

Table 2: Average annual violation days for different $\beta$’s.

<table>
<thead>
<tr>
<th>dispatch</th>
<th>Ave annual violation days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current economic dispatch</td>
<td>114.2</td>
</tr>
<tr>
<td>Balanced dispatch ($\beta = 50%$)</td>
<td>60.8</td>
</tr>
<tr>
<td>Balanced dispatch ($\beta = 85%$)</td>
<td>7.2</td>
</tr>
<tr>
<td>Balanced dispatch ($\beta = 115%$)</td>
<td>0.2</td>
</tr>
<tr>
<td>Balanced dispatch ($\beta = 150%$)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

If we go a step further and apply the ADi Algorithm (from Section 4) to Balanced Dispatch ("Augmented Balanced Dispatch"), we are able to completely resolve the Inertia issue at minimal additional cost. In other words, by replacing current Economic Dispatch with our Balanced Dispatch, not only is the average generation cost largely reduced, but as an added bonus, the system inertia is also maintained at almost no additional cost.

Fig. 12 shows the average generation costs for four simulation settings we run in this paper:

Set. 1: current Economic Dispatch with no action to fix major Inertia violation (blue curve);
Set. 2: Economic Dispatch with our Augmented Dispatch for Inertia (ADi) Algorithm, paying a premium to select Convs to maintain system Inertia (purple curve);
Set. 3: our Balanced Dispatch with no other action, which produces only minor Inertia violations (olive curve); and
Set. 4: our Balanced Dispatch with our ADi Algorithm that prevents all Inertia violations (gold curve).

Economic Dispatch without Augmented Usage (Set. 1) has the highest average generation cost and also results in the least stable Inertia condition. Without external action, it often does not prevent Inertia violations.

In comparing the purple and olive curves (Set. 2 & 3), we can see that even though Economic Dispatch with Augmented Usage produces a lower average cost than without it, the cost is still much higher than if we planned ahead and saved some Convs using Balanced Dispatch.

The olive and gold curves (Set. 3 & 4) almost overlap, indicating the extra cost to ensure Inertia for Balanced Dispatch at $\beta = 85\%$ is negligible. With only a few violation days to fix under Balanced Dispatch, this is in alignment with the results in Table 2 and Fig. 11. The detailed plant progression for Set. 4 is displayed in Fig. 13. The Augmented Usage (violet bars between the orange and blue bars) is barely visible. By comparing to the cost of Balanced Dispatch without Augmented Usage in Fig. 12, we can see that there is almost no additional cost for applying the ADi Algorithm to Balanced Dispatch at $\beta = 85\%$. It proves the ADi (Alg. 1) used with current Economic Dispatch benefits the system with both cheaper long-term costs and fixed grid Inertia. Also, Balanced Dispatch (Alg. 2) alone near-optimally ensures both cheap long-term cost and consistent grid Inertial resilience, but together Alg. 1 and Alg. 2 (Augmented Balanced Dispatch) produce the best results.

Figure 13: Detailed plant usage at $\beta = 85\%$ with Augmented Balanced Dispatch.

In summary, the ADi method leads to a visible improvement over the current Economic Dispatch. Balanced Dispatch though goes even further, achieving the lowest average long-term generation cost of all the methods while at the same time largely improving grid resilience. Furthermore, the ADi method coupled with Balanced Dispatch, yields yet another improvement as the system is able to
both minimize its generation cost and guarantee itself stable system Inertia.

In future work, it would be interesting to investigate the effects of large solar penetration, as well as conduct simulations with datasets from other regions.

8 CONCLUSION

This paper has investigated the consequences of current Economic Dispatch on the long-term generation cost and on grid Inertia stability. Current Economic Dispatch largely operates under the assumption that a free market will discover the best generation mix at the best price. We have shown this base assumption to be incorrect with the current Economic Dispatch method actually unnecessarily leading to both notably escalated generation cost and severely inadequate grid Inertia stability.

On the other hand, our proposed Balanced Dispatch and ADI solutions have demonstrated that a more systematic approach via a healthy generation mix is very effective in achieving both a notable reduction in long-term generation cost and a systematic improvement in grid resilience.

ACKNOWLEDGMENTS

Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing, UC Berkeley. We gratefully acknowledge the financial support from NSF grants CCF-1733832, CCF-1331863, and CCF-1350823.

We thank the reviewers for their multiple insightful comments, one of which was to define Inertia as a product that resources bid for, similarly to capacity markets so as to avoid seeming favoritism of conventional resources. We leave this interesting idea and implementation for future work, as well as a comparison of our proposed approach with the use of capacity markets.

REFERENCES


A INERTIA CONSTANTS

ERCOT provides inertia constants for different generation types. Table 3 indicates that Combined-Cycle Gas, Combustion Gas Turbines and Nuclear have larger inertia constants than Coal, Gas Steam and Hydro; while non-synchronous, inverter-based resources like Wind and Solar do not contribute at all to 'synchronous' Inertia [18].
Table 3: Inertia constant ranges by resource type [18].

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>Inertia Constant Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear</td>
<td>3.8–4.34</td>
</tr>
<tr>
<td>Coal</td>
<td>2.9–4.5</td>
</tr>
<tr>
<td>Combustion Turbine</td>
<td>1–12.5</td>
</tr>
<tr>
<td>Gas Steam</td>
<td>1–5.4</td>
</tr>
<tr>
<td>Combined Cycle</td>
<td>1.1–9</td>
</tr>
<tr>
<td>Hydro</td>
<td>2–3</td>
</tr>
<tr>
<td>Wind</td>
<td>0</td>
</tr>
<tr>
<td>Solar</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: The installed capacity of each generator type in ERCOT for the 2019 baseline scenario [7, 19].

<table>
<thead>
<tr>
<th>Generator Type</th>
<th>Installed Capacities (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear</td>
<td>4,960</td>
</tr>
<tr>
<td>Coal</td>
<td>14,225</td>
</tr>
<tr>
<td>Gas Comb. Turbine</td>
<td>5,901</td>
</tr>
<tr>
<td>Gas Steam</td>
<td>6,200</td>
</tr>
<tr>
<td>Gas Combined-Cycle</td>
<td>32,572</td>
</tr>
<tr>
<td>Gas Engine</td>
<td>700</td>
</tr>
<tr>
<td>Hydropower</td>
<td>557</td>
</tr>
<tr>
<td>Biomass</td>
<td>186</td>
</tr>
<tr>
<td>Wind (inland)</td>
<td>19,228</td>
</tr>
<tr>
<td>Wind (coastal)</td>
<td>2,821</td>
</tr>
<tr>
<td>Solar</td>
<td>1,861</td>
</tr>
</tbody>
</table>

Table 4 shows the baseline energy mix of ERCOT in 2019. We added the capacities of newly installed plants in 2018 and 2019 from ERCOT [19] to the already summarized 2017 baseline statistics [7]. As one can see, most gas generators are combined-cycle.

Lazard conducted an LCOE (levelized cost of energy) analysis in 2019. Fig. 14 shows a table from this report comparing the LCOE of newly-built Renewable generation (onshore Wind and utility Solar) to the marginal cost of existing Conventional generation (Coal and Nuclear) [9]. LCOE Tax Subsidies, LCOE Low End and LCOE High End in this Lazard report were also utilized as a basis to derive the prices in Table 1. Note that: i) from writer’s inside industry knowledge, ERCOT plants are usually on average priced a bit cheaper than the rest of the country and ii) for the most part, plant prices in Table 1 are set around $5 above their marginal costs, due to plants need to make some profit to justify staying in business. Example: We derive the Coal cost range in Table 1 by taking the Lazard number of $26–$41 from Fig. 14 and tighten the band to get rid of outliers, resulting in $29–$37. Then we drop the top 1/4 due to ERCOT being overall a bit cheaper than other US regions, which comes to $29–$35. Lastly, we add $5 assuming that generators would want to make some profit above their marginal cost. This gives us $34–$39. For Nuclear, following a similar method, we start with $27–$31 and add $5 above marginal cost, coming to $32–$36.

We create two empirical distributions for ‘inland’ Wind and Coastal Wind capacity factors to reflect the variable capacity of wind generation. Both distributions are generalized from the Wind factor distribution of a wind farm located in North Texas (purple curve in Fig. 15). For the inland Wind factor, we use a mean of 0.38 and for coastal a mean of 0.42.

For Solar generation, the capacity factor follows a truncated normal distribution centered at 0.5 with a standard deviation of 0.2. Additionally, all capacity factors are bounded below by 0.1 and above by 0.75. This being due to the fact that in large geographical areas (i.e. ERCOT), cumulative Wind or Solar production never stops, nor does it ever reach 100% on all plants at the same time.

For demand, we use ERCOT’s monthly peak demand as the upper bound and since ERCOT does not specify lower demands, we assume the lower bound of the demand is at 80% of the provided average monthly demands [20]. We justify this with the fact (based on ERCOT 24 hour demand data) that within a given day, the minimum demand is approximately 60% of the same day’s peak on
average throughout the year and second, that average monthly demand is relatively centered between the peak and lower values. Further, with no ERCOT demand data after year 2029, we assume the demand increase to be at the rate of 1% for each year after 2029. We select the later years growth value from the expectation that the current 2+% growth rate, mainly from industry (largely oil development) and population growth, is unsustainable and second, from the expectation that energy efficiency within modern construction practices will also help subdue future demand growth.

E  OPTIMAL CHOICE OF $\beta$

To find the optimal threshold of Balanced Dispatch, we calculate the average generation cost (over the entire time horizon $T$) for different values of $\beta$. For our simulation setting, the result is shown in Fig. 16. As one can see, there is a significant reduction in the average generation cost compared to the standard Economic Dispatch when $\beta$ is chosen appropriately. In general, the optimal value of $\beta$ can be efficiently found by binary search.

F  MORE PLOTS ON PLANT USAGE PROGRESSION

The detailed plant use with Balanced Dispatch for various additional $\beta$’s is shown in Fig. 17-19. As $\beta$ increases from 0% to 150% (in Fig. 5, 17, 11 (optimal), 18, 19), Peaker use is reduced, mainly due to the lower use of Renewables, and Convs use increases. The red, yellow, blue, orange and green bars in the Figures are Peakers, new Convs, initial Convs, new Renewables and initial Renewables respectively.

While these larger $\beta$ scenarios (> 100%) in Fig. 18-19 save more Convs and are more resistant to variability factors, they fail to utilize a healthy balance of cheaper Renewables and therefore result in a higher average generation cost than the optimal $\beta = 85\%$.

It is also worth noting in Fig. 17-19 that: i) ‘less active’ Convs (the light blue & light yellow bars) are much more common in higher $\beta$’s (> 100%) than in lower ones (< 80%) and ii) Renewables are extremely and unnecessarily suppressed in higher $\beta$’s (> 100%).