Delft University of Technology

Flexibility Framework With Recovery Guarantees for Aggregated Energy Storage Devices

Evans, Michael; Tindemans, Simon H.; Angeli, David

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Flexibility Framework With Recovery Guarantees for Aggregated Energy Storage Devices

Michael P. Evans⬤, Simon H. Tindemans⬤, *Member, IEEE*, and David Angeli⬤, *Fellow, IEEE*

Abstract—This paper proposes a framework for the procurement of flexibility reserve from aggregated storage fleets. It allows for arbitrary tree structures of aggregation hierarchy, as well as easily implementable disaggregation via broadcast dispatch. By coupling discharge and recovery modes, the proposed framework enables full-cycle capacity to be procured ahead of real time, with guaranteed recovery and exact accounting for losses. The set of feasible discharging requests is exactly encoded, so that there is no reduction in the ability to meet discharging signals, and recovery capabilities are parameterized as a single virtual battery. Included in this paper is a numerical demonstration of the construction of the constituent curves of the framework and the approach is also benchmarked against relevant alternatives.

Index Terms—Flexibility, aggregation, virtual power plant, energy storage systems, optimal control, ancillary service, virtual battery.

I. INTRODUCTION

A. Background

RECENT years have seen an increased proliferation of renewable energy sources onto electricity networks, leading to increased difficulty in the task of balancing supply and demand. Energy storage units offer significant potential in addressing this imbalance. For example, National Grid, the Electricity System Operator, predicts that between 28 and 75 GW of electricity storage (including vehicle-to-grid) capacity will be used to balance the British network by 2050 [1]. Prior literature on balancing services has covered multiple classes of storage device, such as EV batteries [2]–[8], diesel generators [9], [10] and home storage units [3], [11]–[15]; and a battery model can even capture the properties of a time-shift in electrical load [16]–[19]. The total capabilities of units that can be modeled as *virtual batteries* is therefore very significant.

A key enabling technology for large numbers of devices to participate in balancing frameworks is their *aggregation*. This combination of individual units into compound entities, often termed Virtual Power Plants, offers a useful abstraction framework: in amalgamating a complex fleet into a single capability representation, one can consider there to be a single service provider. This leads to ease of interaction with the service recipient, as well as a reduction in the direct communication infrastructure connecting individual device owners and the central dispatcher.

Multiple authors have considered the aggregation of energy storage units [4]–[7], [17], [20]. The predominant form of aggregation has been the summation of power and energy limits across devices, as exemplified by [5]. While this functions as intended for homogeneous fleets, we showed in [21] that such a representation only serves to outer-bound the true flexibility limits of a heterogeneous device pool. The summation-aggregation approach was improved upon in [4], where Vandael *et al.* presented a receding horizon version of the summed parameter model, with the addition of necessary constraints on total charge profiles; this same tightening of the aggregate energy constraints was then applied in [6], [7]. However, the retention of a summed power limit allowed the fleet to run at full power up to the point of full charge (equivalent to energy depletion in a discharging setting). When such an aggregation approach is utilized to compose constraints on an optimization problem, this can still lead to the allocation of infeasible requests. Appino *et al.* noted this shortfall of summation based aggregation in [20]. While our approach is to directly rectify this problem, they instead imposed conditions so that a summation *was* able to exactly capture the flexibility limits of the fleet. In doing so, these conditions necessarily restrict the aggregate flexibility, however, which may result in sub-optimal allocation when a feasible optimum is placed outside of the considered region. By contrast, the control approach presented in [21] was able to exactly capture aggregate discharging capabilities. There, we presented a transform which exactly encodes fleet flexibility, under the restriction to unidirectional operation and a lack of cross-charging, termed the *E-p* transform. This is the predominant tool that we will use in our analysis here, applied to coupled discharge-recharge operation. An equivalent fleet

Michael P. Evans was with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BX, U.K. He is now with Octopus Energy, London W1D 1NN, U.K. (e-mail: michael.p.evans21@gmail.com).

Simon H. Tindemans was with the Data-Centric Engineering Programme, Alan Turing Institute, London NW1 2DB, U.K. He is now with the Department of Electrical Sustainable Energy, TU Delft, 2628 CD Delft, The Netherlands (e-mail: s.h.tindemans@tudelft.nl).

David Angeli is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BX, U.K., and also with the Department of Information Engineering, University of Florence, 50121 Florence, Italy (e-mail: d.angeli@imperial.ac.uk).

flexibility measure was presented by Cruise and Zachary in [22].

In addition to the aforementioned benefits, aggregation (either directly or in stages) also enables an anonymization of individual unit owners. Moreover, where device availability is stochastic, the aggregation of units leads to a statistical smoothing of the overall behavior, i.e., a robustification of the net output. For example, chance constraints can be applied to achieve an arbitrary risk level (as in [23]) only on aggregate.

Network operators are increasingly offering frameworks for balancing services, procured through ancillary service markets (e.g., [24]–[27]). These markets are generally accessed by aggregators, entities which combine the capabilities of portfolios of heterogeneous units into compact delivery offers; grid operators are increasingly developing software platforms to enable scalable and automated participation via this route-to-market (see [28] for an example). Once these markets exist, they encourage the formation of complex ecosystems in which entities of various types and sizes exchange arbitrary ancillary service provisions with one another. In this regard, a desirable feature in an aggregation scheme is the ability to abstract the capabilities of a fleet in such a way that aggregates themselves can be combined into larger aggregates. This property enables *hierarchical* aggregation, which improves scalability and underpins combination and re-packaging of services.

Examples of aggregation of heterogeneous resources exist for various load types. For example, for thermostatically controlled loads, aggregate capability representations were presented in [29], [30]. Zhao *et al.* [31] present an hierarchical aggregation method based on homothets of a prototype polytope (which encodes a standard measure of flexibility), where inner (sufficient) and outer (necessary) flexibility bounds are calculated and aggregated. Uncertainty in device parameters was introduced in [32], using a chance-constrained formulation. Homothet-based geometric addition was also used for active-reactive power (*P-Q*) flexibility in [33]. Müller *et al.* introduced zonotopes as an alternative encoding of resource flexibility polytopes, based on an inner approximation (feasibility) [34]. Zonotopes also feature straightforward aggregation, but are based on a standard set of generators instead of (homothets of) a prototype polytope. Finally, Ulbig and Andersson proposed a hierarchical aggregation scheme that features ramp rate constraints in addition to power and energy constraints [35], but feasibility is no longer guaranteed after aggregation. All these methods for aggregating heterogeneous flexibility resources sacrifice either flexibility or feasibility guarantees for tractability. A qualitatively different approach is taken in [36], where a real-time aggregate flexibility metric is learned within a reinforcement learning framework, which is suitable for real-time control applications, but less so for markets and scheduling, as the product is not easily quantified ahead of time. Tsaousoglou *et al.* [37], considering the problem of optimal EV charging, similarly forego an explicit representation of flexibility and focus on the distributed dispatch problem.

A particular challenge when deploying demand response services is the need to schedule recovery (or recharging in the case of batteries), potentially under a strict deadline.
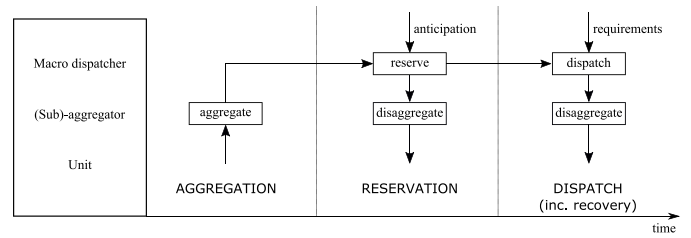

Fig. 1. The separation in time of aggregation, reserve scheduling and dispatch according to the presented framework.

O'Connell *et al.* [17] proposed asymmetric block offers as a means of co-optimizing response and rebound in a single operation. By packaging response and rebound (or equivalently discharging and recharging) into a single offer, they guarantee full recovery at a specified time, thus reducing risk exposure.

B. Contributions

This paper introduces a framework for hierarchical aggregation of heterogeneous storage fleets that guarantees energy recovery: the *discharge-loss-recovery* (DLR) framework. This addresses the same challenge as [17], but extends the approach in two important ways: first, in place of a single virtual battery, the DLR representation precisely encodes the aggregate capabilities of heterogeneous storage units. Second, instead of a discrete set of offers, a continuous set of capabilities is communicated. This framework enables ahead-of-time scheduling without sacrificing achievability of recovery. As such, it enables a macro dispatcher (e.g., a system operator) to reserve flexibility in a consistent way, bounding the operational risk (and reducing associated costs) as compared to transmitting emergency requests in real time. This results in the three-stage process shown in Figure 1. At present, there are three major use-cases for ahead-of-time reservation of flexibility: 1) grid balancing, contracted by the TSO; 2) portfolio balancing, contracted by balance responsible parties; and 3) congestion management, contracted by DSOs.

The DLR framework encodes the flexibility of storage units for the (sub-)aggregators and the macro-dispatcher, and incorporates dispatching of these units. It offers the following properties.

- Accounting for heterogeneity of storage units and round-trip losses.
- Exact encoding of aggregate discharge flexibility.
- After ensuring maximum discharge flexibility, *a priori* determination of recovery requirements (energy and minimum recharge time).
- Hierarchical aggregation of flexibility and disaggregation via broadcast dispatch.

Reflecting the predominant use of storage under supply-shortfall conditions today, we focus on discharging requests followed by recharging recovery. The presented technique can then be applied to the opposite case via a straightforward inversion of the coordinate system used.

The contributions of this paper are as follows.

- We present a framework for coupling response and recovery capabilities that accounts for losses and asymmetry in discharging and recharging rates. This is critical to

flexibility reservation involving energy-limited resources and as such improves upon the unidirectional reservation-plus-dispatch that is used by system operators for balancing today. Not only would implementing our framework ensure successful recovery, but it would also allow a system operator to control said recovery and regulate the system in doing so.

- Using system balancing data, we investigate the qualitative benefits that application of our framework would bring over the purchase of multiple block offers of fixed power across the delivery window.
- We provide closed-form equations for the aggregation of capability packets from heterogeneous device sub-fleets and describe the accompanying broadcast dispatch; together these make the approach scalable in the sense described in [4].
- We qualitatively benchmark our approach against the state-of-the-art.
- We discuss how the framework can be applied to model demand response.
- We discuss how the framework can be applied in stochastic settings.
- We demonstrate construction of a range of DLR packets and, in doing so, show how insights can be gained from these capability encodings.

II. PROBLEM DESCRIPTION

A. Mathematical Description

We denote the constituent heterogeneous units of the fleet as D_i, $i = 1, \ldots, n$, each with extractable energy $e_i(t)$ subject to the assumed physical constraint $0 \leq e_i(t) \leq \bar{e}_i$. As our focus is on the ability of a fleet to counter supply-shortfall conditions, we frame the problem based on discharging operation, choosing the sign convention that discharging power outputs are positive. We choose as our control input $u_i(t)$ the power extracted from each device, measured externally so as to take into account any inefficiencies present during discharging operation. This then leads to the following integrator dynamics:

$$\dot{e}_i(t) = \begin{cases} -u_i(t), & \text{if } u_i(t) \geq 0 \\ -\eta_i u_i(t), & \text{otherwise,} \end{cases} \tag{1}$$

in which $\eta_i \in (0, 1]$ denotes the round-trip efficiency of device D_i.

We set our problem within the paradigm of response-recovery operation, i.e., where a fleet responds by discharging available units in a nominal delivery phase, followed by a recharging phase in which the initial energy state is recovered. Initially, we focus on the discharging mode only, defining the concepts we will use for this setting specifically, before extending the same concepts to recharging mode through necessary changes in coordinates in Section II-E. To this end, we initially set the power limits of each device as $u_i(t) \in [0, \bar{p}_i]$. We assume a lack of cross-charging, by which we mean the use of available headroom to redistribute

energy among devices.[1] We choose as our state variable the *time-to-go*, defined for each device as $x_i(t) \doteq e_i(t)/\bar{p}_i$. We then stack state, input and maximum power values across devices as follows:

$$x(t) \doteq \begin{bmatrix} x_1(t) & \ldots & x_n(t) \end{bmatrix}^T, \tag{2}$$

$$u(t) \doteq \begin{bmatrix} u_1(t) & \ldots & u_n(t) \end{bmatrix}^T, \tag{3}$$

$$\bar{p} \doteq \begin{bmatrix} \bar{p}_1 & \ldots & \bar{p}_n \end{bmatrix}^T, \tag{4}$$

allowing us to rewrite the discharging dynamics in matrix form as $\dot{x}(t) = -P^{-1}u(t)$, in which $P \doteq \text{diag}(\bar{p})$. We also form the product set of the power constraints, $\mathcal{U}_{\bar{p}} \doteq [0, \bar{p}_1] \times [0, \bar{p}_2] \times \cdots \times [0, \bar{p}_n]$, so that our discharging constraints can be compactly written as $u(t) \in \mathcal{U}_{\bar{p}}$ and $x(t) \geq 0$.

B. Feasibility and Flexibility

The intention of this paper is to present a framework through which a macro dispatcher can procure flexibility reserve, ahead of real-time dispatch of the reserved capabilities. This generalize the concept of a two-stage scheduling-plus-dispatch routine (see [21] for further details), since the reserve circumscribes the set of admissible schedules. For consistency, then, it is imperative that the considered framework encodes those schedules that are *feasible*, by which we mean that they do not violate any physical device constraints. Initially, we focus our attention on pure discharging requests, which we allocate as positive by convention and denote[2] by $P^d \colon [0, +\infty) \mapsto [0, +\infty)$. We then characterize the set of feasible requests and define flexibility as follows:

Definition 1: The **feasible set** of discharging requests, for a system with maximum power vector $\bar{p}$ and initial state $x = x(0)$, is defined as

$$\mathcal{S}_{\bar{p},x} \doteq \Big\{ P^d(\cdot) \colon \exists u(\cdot), \; z(\cdot) \colon \forall t \geq 0, \; 1^T u(t) = P^d(t),$$
$$u(t) \in \mathcal{U}_{\bar{p}}, \; \dot{z}(t) = -P^{-1}u(t), \; z(0) = x, \; z(t) \geq 0 \Big\}.$$

Definition 2: The **flexibility** of an energy resource is an encoding of the feasible set that has the following properties:
1) A numerical representation of admissible requests
2) An accompanying dispatch routine to realize admissible requests
3) Guaranteed feasibility of each admissible request within *a priori* bounds

We say that flexibility is maximized if the bounds on admissible signals exactly encode the feasible set, thus guaranteeing feasibility without restricting the operational envelope.

C. The E-p Transform

We presented an example of flexibility maximization in [21], precipitated through the restriction to pure discharging operation, termed the *E-p* transform. We will utilize this here and so make the following reproductions:

---

[1] It is shown in [38] that cross-charging does not increase the feasible set for fully charged storage devices serving a unimodal signal. Moreover, cross-charging of devices with less than 100% efficiency necessarily results in a loss of energy, which is often undesirable.

[2] In the interest of clarity, some of the notation used here will differ from that of our prior work [21], [23], [39], [40].
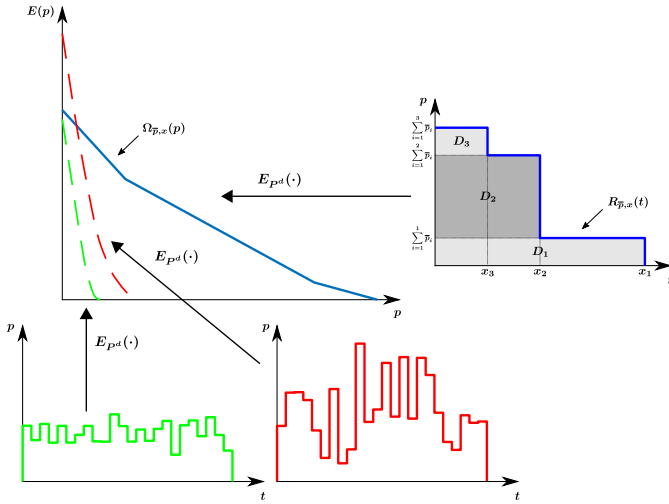
Fig. 2. An example use of the *E-p* transform as a binary feasibility check, implementing Property 1. The capacity curve is formed based on fleet parameters, and received requests are transformed into *E-p* space; at which point the binary feasibility check can be made by eye. The green curve is found to be feasible, while the red curve is found to be infeasible.

Definition 3: Given a discharging power request $P^d(\cdot)$, we define its *E-p* transform as the following function:

$$E_{P^d}(p) \doteq \int_0^\infty \max\left\{P^d(t) - p, 0\right\} dt,$$

interpretable as the energy required above any given power rating, p. The *E-p* transform is convex and monotone.

Definition 4: We define the discharging **capacity** of a system to be the *E-p* transform of the worst-case request that it can meet, i.e.,

$$\Omega_{\overline{p},x}(p) \doteq E_{R_{\overline{p},x}}(p),$$

with $R_{\overline{p},x}(t) \doteq \sum_{i=1}^n \overline{p}_i[H(t)-H(t-x_i)]$, in which $H(\cdot)$ denotes the Heaviside step function.

Property 1: A discharging request $P^d(\cdot)$ is feasible if, and only if, its *E-p* transform is dominated by the capacity of the system, i.e., $P^d(\cdot) \in \mathcal{S}_{\overline{p},x} \iff E_{P^d}(p) \leq \Omega_{\overline{p},x}(p) \ \forall p$.

A graphical example of the use of the *E-p* transform, via Property 1, as a binary feasibility check can be seen in Figure 2. We reiterate that, under enforced conditions, the capacity curve exactly encodes the feasible set. We can therefore formally define the corresponding flexibility as follows:

Definition 5: The **pure discharging flexibility** of a fleet with maximum power vector $\overline{p}$ in state x is defined as

$$\mathcal{F}_{\overline{p},x} \doteq \left\{(p, E)\colon p \geq 0, \ E \leq \Omega_{\overline{p},x}(p)\right\}$$
$$= \left\{(p, E_{P^d}(p))\colon p \geq 0, P^d(\cdot) \in \mathcal{S}_{\overline{p},x}\right\}.$$

It also has the following property that will be useful in our analysis:

Property 2: The complement to the discharging fleet flexibility, $\mathcal{F}_{\overline{p},x}^{\complement} = ([0, +\infty) \times [0, +\infty)) \setminus \mathcal{F}_{\overline{p},x}$, is a convex set.

D. Default Dispatch Policy

In [39], we presented a policy and showed it to maximize the flexibility of the fleet at all future time instants. We then

presented a discrete time equivalent to this in [40], which allocates a piecewise constant control input to match a piecewise constant request. In our framework, we will utilize the latter for default dispatch and therefore reproduce it here as follows. We restrict ourselves to a constant dispatch period Δt, envisioning that this will be most relevant in practical settings. Given a request that takes a constant value P^d across the time interval, a target state value $\hat{z}$ is found as

$$\hat{z} : \sum_{i=1}^n \overline{p}_i \max\left\{0, \min\left\{x_i - \hat{z}, \Delta t\right\}\right\} = P^d \Delta t, \quad (5a)$$

then the input is constructed according to

$$u_i = \overline{p}_i \cdot \max\left\{0, \min\left\{\frac{x_i - \hat{z}}{\Delta t}, 1\right\}\right\}, \quad (5b)$$

resulting in time-to-go updates

$$x_i \leftarrow x_i - \frac{u_i \Delta t}{\overline{p}_i}. \quad (5c)$$

Note that the algorithm is applied at each time step, so explicit time-dependence is omitted. Also, unlike in [40], we need only consider feasible requests, which leads to equality in the energy balance of (5a).

E. Symmetry to Recharging Operation

Up to this point, we have considered pure discharge operation. We will now extend this to the recharging of devices between shortfall events. We denote a recharging request as $P^r : [0, +\infty) \mapsto (-\infty, 0]$. Analogous to time-to-go, we utilize the *time-to-charge*, defined as follows:

$$y_i(t) \doteq \frac{\overline{e}_i - e_i(t)}{\eta_i \underline{p}_i}, \quad (6)$$

in which $\underline{p}_i$ denotes the (magnitude of the) maximum charge rate of device D_i. Equivalently, this can be obtained from a time-to-go value via the following change of (instantaneous) variables:

$$y_i = \frac{\overline{p}_i(\overline{x}_i - x_i)}{\eta_i \underline{p}_i}, \quad (7)$$

in which $\overline{x}_i \doteq \overline{e}_i/\overline{p}_i$ is the maximum time-to-go value of device D_i. When considering pure recovery operation, then, we are able to perform the conversions $x_i \to y_i$, $P^d \to -P^r$, $\overline{p}_i \to \underline{p}_i$ and $u_i \to -u_i^r$, where u^r denotes the (negative) control input while recharging, and the equivalent results hold.

Just as the scheduling from an arbitrary state can be considered in either direction, the corresponding default dispatch policy will be given by either (5) or its equivalent in terms of time-to-charge. The user is alternatively able to compose the recharging policy based directly on discharging parameters as follows. Given a request that takes a constant value P^r across the time interval Δt, a target (recharging) state value $\hat{y}$ is found as

$$\hat{y} : \sum_{i=1}^n \underline{p}_i \max\left\{0, \min\left\{\frac{\overline{p}_i(\overline{x}_i - x_i)}{\eta_i \underline{p}_i} - \hat{y}, \Delta t\right\}\right\} = -P^r \Delta t, \quad (8a)$$

then the input is constructed according to

$$u_i^r = -\underline{p}_i \cdot \max\left\{0, \min\left\{\frac{\overline{p}_i(\overline{x}_i - x_i) - \eta_i \underline{p}_i \hat{y}}{\eta_i \underline{p}_i \Delta t}, 1\right\}\right\}, \quad (8b)$$

resulting in time-to-go updates

$$x_i \leftarrow x_i - \frac{\eta_i u_i \Delta t}{\overline{p}_i}. \quad (8c)$$

III. The DLR Framework

A. Sequential Versus Integrated Scheduling

Our intention is to construct a single flexibility representation that encodes the feasible set (under no cross-charging) for the full discharge-recovery cycle. This then allows for scheduling of the complete cycle ahead of real-time.

We restrict our analysis to requests that are positive or zero (discharging) over some period, followed by negative or zero (recharging) over another period, such that full recharging of all devices is achieved; this cycle can then be repeated. We do not explicitly limit the duration of either period, but such a constraint can be communicated as an additional parameter. In line with the previous section, we shall denote a discharge-recovery request as $P^d \oplus P^r(\cdot)$, in which the $\oplus$ operator denotes the concatenation of two signals.

The results in Section II provide the tools to implement a two-stage solution to this problem. First, the discharging capacity curve (which, to emphasize the mode, we denote as $\Omega^d(\cdot) \doteq \Omega_{\overline{p},x}(\cdot)$) is used to determine feasibility of a discharge request $P^d(\cdot)$, and the dispatch policy (5) used to realize it. Then, the procedure is repeated for the recharge request $P^r(\cdot)$. However, the *recharge capacity curve* $\Omega^r(\cdot)$ and, importantly, the total amount of energy required to recharge while accounting for losses, depend on the terminal state-of-charge of the storage units after the discharge operation, which in turn depends on (the *E-p* transform of) the discharge signal $P^d(\cdot)$. This dependency hinders the formulation of a compact flexibility representation that incorporates recharging.

B. The DLR Flexibility Representation

In this section, we introduce a representation of the set of feasible integrated discharge-recovery schedules that has the following properties.
- It maximizes the discharging flexibility, so that the set of considered discharging signals exactly matches the feasible set.
- It exactly accounts for round-trip losses in the recharging phase. This is a crucial requirement for any aggregation method used to produce constraints on an optimization problem.
- It returns recharge parameters in the form of a simple virtual battery model, characterized by recharge energy and minimum time-to-charge (equivalently recharge power and energy ratings).

As mentioned in the previous section, the recharge requirements depend on the terminal state of the discharge procedure. In order to simplify the description of recharge flexibility, we introduce a one-dimensional parameterization of the discharging flexibility as a function of the total energy requirement E^d,

in such a way that discharging flexibility is not sacrificed and an exact encoding of the recharge requirements is possible. This representation reduces the feasible set of recharging patterns, in exchange for a compact representation that facilitates pre-scheduling of recharging.

The proposed *discharge-loss-recovery time* (DLR) representation consists of three 1-parameter functions:
- D) The *discharge* capacity curve $\Omega^d(\cdot)$ (Definition 4);
- L) The *loss* function, describing the energy required to recharge, E^r, as a function of the reserved discharge energy E^d;
- R) The *recovery time* function, describing the minimum time, y^*, required to recharge the reserved discharge energy E^d.

Together, the loss and recovery time functions define the E^d-dependent parameters of a single virtual battery for recharging. For ease of aggregation and broadcast, we choose to parameterize the loss and recovery time functions by x^*, the minimum time required to discharge the reserved discharge energy E^d, in place of E^d itself. This is detailed later in the section.

For a single battery, the three constituent curves of the DLR framework are composed via the following change in coordinates from the (scalar valued) parameter tuple $(\overline{p}, x, \underline{p}, \eta)$:

$$\Omega_{\text{single}}^d(p) = x(\overline{p} - p), \quad (9a)$$

$$E_{\text{single}}^r(x^*) = \frac{\overline{p}x^*}{\eta}, \quad (9b)$$

$$y_{\text{single}}^*(x^*) = \frac{\overline{p}x^*}{\eta \underline{p}}, \quad (9c)$$

where E_{single}^r and y_{single}^* jointly define the recharging unit. Example curves for a single unit can be seen in Figure 3 (upper panel).

The equivalent curves for a heterogeneous fleet can be seen in Figure 3 (lower panels). Details on the construction of the DLR-curves for arbitrarily sized fleets are provided in the remainder of this section.

C. Discharging Capacity and Guaranteed Final State

The first element of the DLR representation, the *discharging capacity* is given by the capacity curve $\Omega^d(\cdot)$, which exactly encodes the feasible set of discharging signals for a heterogeneous fleet.

Given a feasible discharge signal $P^d(\cdot)$ and an initial state, the dispatch strategy (5) determines the terminal state-of-charge of each storage unit. In order to reduce the space of possible outcomes without limiting feasibility, we construct a *maximum-flexibility truncated fleet* that limits the time-to-go of each storage unit. This truncated fleet is parameterized by E^d and will be exactly emptied during discharge operation and refilled during recovery, *without* restricting the feasible set conditioned on the discharged energy E^d. For a given fleet of power rating $\overline{p}$ in state x, we define

$$\tilde{x}_i = \min\{x_i, x^*\}, \quad i = 1, \ldots, n, \quad (10a)$$

in which the (aggregate) minimum discharge time x^* is defined according to

$$x^*(E^d) : \sum_{i=1}^{n} \overline{p}_i \min\{x_i, x^*\} = E^d. \quad (10b)$$
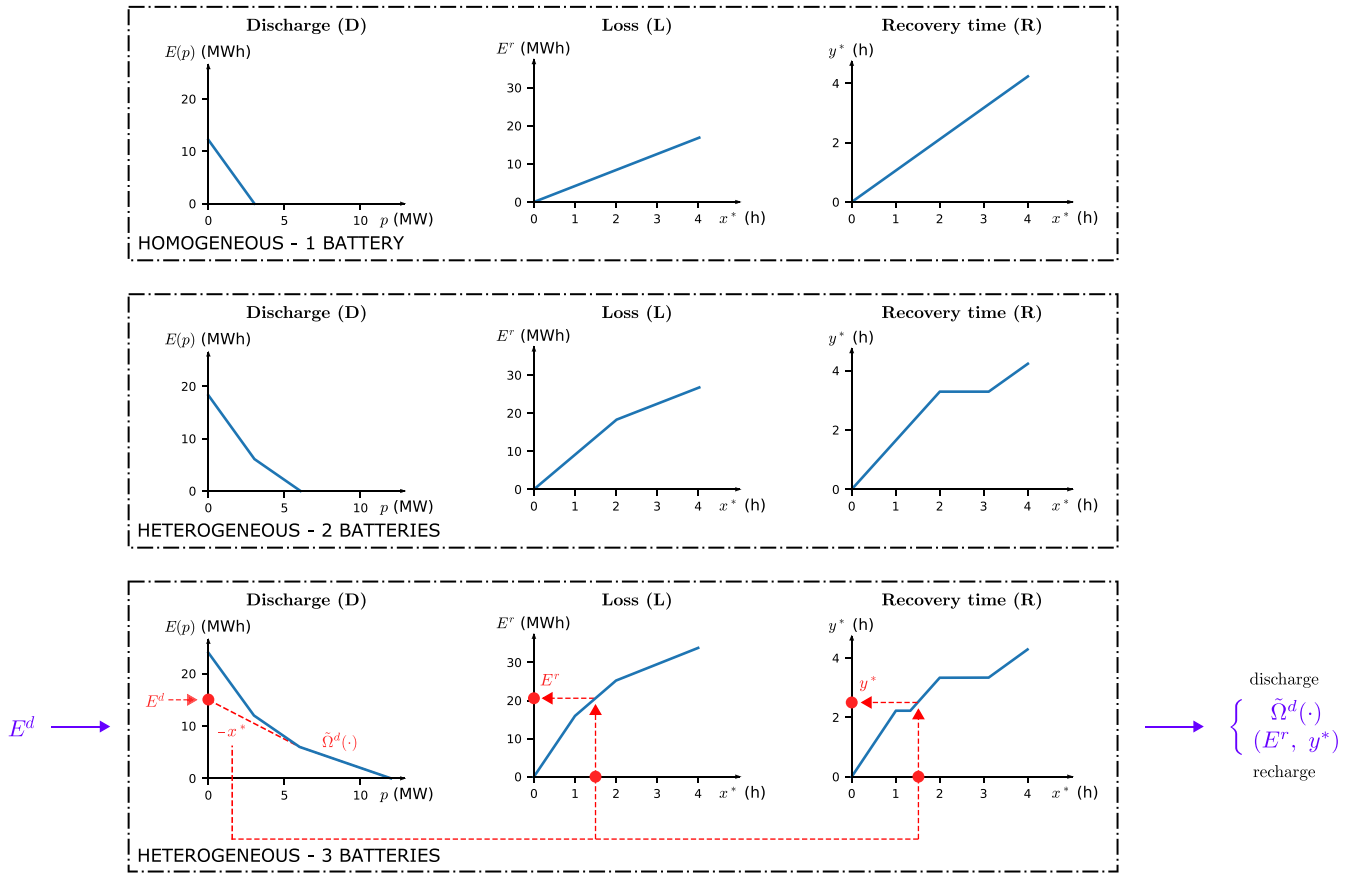
Fig. 3. The DLR curves for a fleets of one (upper panel) two (center panel) and three (lower panel) batteries. In the lower panel, red construction lines show the conversion of a discharge energy E^d into the corresponding discharging and recharging flexibilities.

This truncated fleet construction satisfies the following result, the proof of which can be found in the Appendix:

Theorem 1: Given a request profile $P^d(\cdot)$ and a fleet with maximum power vector $\bar{p}$ in state x,

$$P^d(\cdot) \in \mathcal{S}_{\bar{p},x} \iff P^d(\cdot) \in \mathcal{S}_{\bar{p},\tilde{x}}, \tag{11}$$

in which $\tilde{x}$ is defined according to (10) with $E^d \geq \int_0^\infty P^d(t)dt$.

Thus, in the knowledge that $P^d(\cdot)$ will have a total energy no greater than E^d, the original fleet can be replaced by a truncated fleet, using (10), and maintain its ability to meet considered signals. Moreover, if the total energy provided *exactly* equals E^d, then the virtual fleet will be emptied and the true fleet is in a known state, regardless of the distribution of power levels in $P^d(\cdot)$. This property gives the sought reduction in recharge complexity.

Theorem 1 also leads to the following Corollary:

Corollary 1: The flexibility of a truncated fleet as in (10) can be found as

$$\mathcal{F}_{\bar{p},\tilde{x}}^{\mathsf{C}} = \mathrm{Conv}\Big(\mathcal{F}_{\bar{p},x}^{\mathsf{C}} \cup \Big\{\big(0, E^d\big)\Big\}\Big), \tag{12}$$

in which $\mathrm{Conv}(\cdot)$ denotes the convex hull operator.

This result gives us a means to construct the capacity curve of the truncated fleet directly from a full capacity curve, without considering constituent devices. Figure 4 demonstrates this graphically where, for simplicity of notation, we define
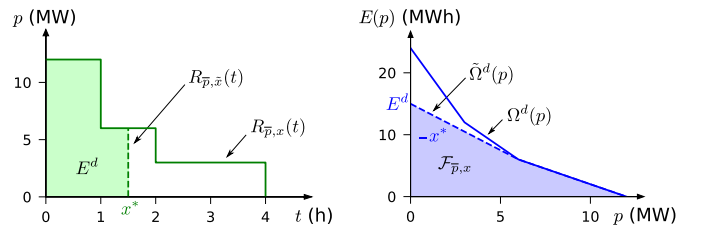

Fig. 4. The correspondence between the worst-case request (left panel) and capacity curve (right panel) for a truncated fleet formed as in (10). The green shaded area is equal to the total requested energy, while the discharging flexibility is shaded in blue.

$\tilde{\Omega}^d(\cdot) \doteq \Omega_{\bar{p},\tilde{x}}(\cdot)$. As can be seen in the figure, the flexibility of a truncated fleet can found via two equivalent methods: 1) time-truncation of the worst-case reference; 2) the convex hull operation of (12) applied directly to the capacity curve. We note that, by construction, the discharge capacity curve of the truncated fleet has the property

$$\left.\frac{d\tilde{\Omega}^d(p)}{dp}\right|_{p=0^+} = -x^*. \tag{13}$$

D. Loss and Recovery Time Functions

Recall that a discharge request with energy E^d will exactly empty an E^d-truncated virtual fleet. The total energy required
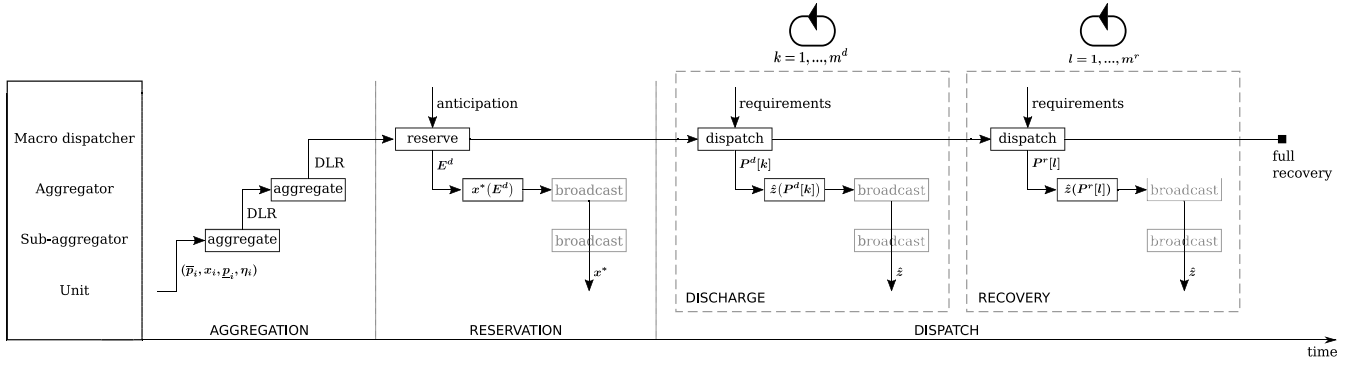
Fig. 5. Aggregation, reservation and dispatch, including the messages passed between entities. $m^d \Delta t$ and $m^r \Delta t$ denote the lengths of the discharging and recharging periods, respectively. Transparent boxes show optional disaggregation steps at one or more sub-aggregator levels.

to recharge is therefore given by

$$E^r(x^*) = \sum_{i=1}^{n} \frac{\overline{p}_i}{\eta_i} \min\{x_i, x^*\}, \qquad (14)$$

with x^* as in (10b). We term this the *loss function* and note that it is piecewise linear, with n sections in general, and concave.[3]

The third and final component of the DLR representation is the *recovery time function*. The initial recharge state corresponds to the truncated fleet being exactly depleted. The fastest recovery is determined by the largest time-to-charge value, which is therefore given by

$$y^*(x^*) \doteq \max_{i \in \{1, \ldots, n\}} \frac{\overline{p}_i}{\eta_i \underline{p}_i} \min\{x_i, x^*\}, \qquad (15)$$

with x^* as in (10b). This function is monotone and piecewise linear, with $2n - 1$ sections in general.

Together, $E^r(x^*)$ and $y^*(x^*)$ describe a virtual storage unit. Acceptable recharging profiles $P^r(\cdot)$ must fully recharge this unit by supplying an amount of energy E^r subject to a maximum power rating E^r/y^*. As it can be seen in Figure 3, then, the constituent DLR curves are sufficient to encode, for a given total discharge energy, coupled discharging and recharging flexibilities. These guarantee feasibility of any allocated request and, as such, the DLR representation should be considered a consistent flexibility metric for integrated discharge-recovery operation. The DLR curves can also be embedded into optimization routines to encode round-trip capabilities.

We also point out here that, in the special case where all devices have the same round-trip efficiency and ratio of charging/discharging rates, implementation of the default discharging policy maximizes *charging* flexibility for a given total energy. Hence, when these parameter conditions apply, full-cycle flexibility is maximized in the sense that the recharge energy and recharge time are minimal for a given discharge energy E^d.

---

[3]The loss and recovery time functions could alternatively be parameterized by E^d in place of x^*. The derivative of the loss function , i.e., $\frac{dE^r}{dE^d}$, would then be a p-weighted average of $1/\eta$ of participating batteries. However, we opt for x^* here for increased convenience of aggregation and broadcast.

IV. AGGREGATION, RESERVATION AND DISPATCH

The framework that we present enables the separation in time of aggregation, flexibility reservation and dispatch, as was shown in Figure 1. Having presented the elements required for implementation of this framework, we here provide details on each of the three constituent time periods, in Figure 5 and the following subsections.

A. Aggregation

The DLR representation is computed on the basis of the fleet parameter tuple $(\overline{p}, x, \underline{p}, \eta)$ (each vector-valued). It is then straightforward to see that DLR representations can be combined: by converting each back to fleet parameters (one virtual device per unique time-to-go value), amalgamating all of these and finally forming a single compound DLR packet. This framework therefore implicitly allows for aggregation in arbitrary hierarchies.

We would, however, like to achieve the combination of DLR packets based directly on the three constituent curves, i.e., without reverting to device parameters. Without loss of generality, we consider the aggregation of two fleets with parameters $(\overline{p}^{(1)}, x^{(1)}, \underline{p}^{(1)}, \eta^{(1)})$ and $(\overline{p}^{(2)}, x^{(2)}, \underline{p}^{(2)}, \eta^{(2)})$. To reduce clutter, we simplify notation by using only the subscript 1 or 2 to summarize the respective quantities. Each of the three DLR curves is treated in one of the following subsections.

1) Discharge Capacity Curve: We initially consider discharge flexibilities, for which we are able to derive the following result (proof in the Appendix):

Theorem 2: The **discharging flexibility** of an aggregate fleet can be computed by the following complementary Minkowski summation:

$$\mathcal{F}_{1+2} = \left(\mathcal{F}_1^{\complement} \oplus \mathcal{F}_2^{\complement}\right)^{\complement}. \qquad (16)$$

This then leads directly to the following result:
Corollary 2: **Capacity curves** can be added as follows:

$$\Omega_{1+2}(p) = \sup\left\{E : (p, E) \notin \mathcal{F}_1^{\complement} \oplus \mathcal{F}_2^{\complement}\right\}. \qquad (17)$$

The combined capacity curve $\Omega_{1+2}(p)$ also defines the fleet truncation parameter $x^*(E^d)$ through (13).

2) Loss Function: A time-to-go value x^* defines all individual device recharge energy needs, regardless of their grouping into fleets. Hence the total energy requirement can be found via a simple pointwise summation, i.e.,

$$E_{1+2}^r(x^*) = E_1^r(x^*) + E_2^r(x^*). \tag{18}$$

3) Recovery Function: The minimum recovery time function is set by the largest time-to-charge value. As such, when two curves are amalgamated, it is the maximum of these values that determines the minimum recharge time, i.e.,

$$y_{1+2}^*(x^*) = \max\{y_1^*(x^*), \ y_2^*(x^*)\}. \tag{19}$$

Finally, we remark that the DLR representation of a fleet can be constructed in this way by aggregating DLR representations of elementary storage units, each given by (9).

B. Flexibility Reservation

Without loss of generality, we assume that the macro dispatcher has a direct communication link to a number of top-level aggregators. Once it has received all DLR capabilities, compound or otherwise, we envision that the macro dispatcher makes a decision as to what flexibility provision to reserve from each; and we place no restrictions on the method through which this is chosen. Once this decision has been made, the actual reservation takes place through the transmission of the amount of reserved energy (one value: E^d) to each top-level aggregator. Each of these then calculates the corresponding x^*-value, which it broadcasts among its fleet. Once reservation has been realized through the allocating of x^* values, a truncated fleet is formed as each unit sets its virtual initial state $\tilde{x}_i = \min\{x_i, x_i^*\}$; this virtual device will be fully discharged and recharged. We point out here that dispatch via the policy (5) ensures participation of all units, and hence so too does the allocation of devices under the proposed framework.

As an extension to this reservation method, the aggregator could in fact amend the truncation level of each device and/or the dispatch strategy, so long as the aggregate constraints were met; in this case it would broadcast a vector x^* of length n instead. However, this is not further investigated here.

C. Broadcast Dispatch

The DLR construction is such that any admissible signal can necessarily be met by the implementation of the policy (5) and its equivalent for recharging. Under this combined policy, each device solely requires the $\hat{z}$ value corresponding to each request level in order to allocate its output. Moreover, this remains true regardless of the grouping of devices into fleets. As a result, the framework that we propose directly enables broadcast dispatch. There is no need to cascade signals down the tree; rather, each top-level aggregator can broadcast a single $\hat{z}$ level and the units will respond accordingly.

D. Benefits of the Combined Framework

We consider this approach to be *scalable* (c.f. [4]), based on its combination of arbitrary aggregation hierarchy and broadcast dispatch. From their construction according to a fleet of

TABLE I
PARAMETERS FOR THE THREE-BATTERY EXAMPLE

Battery #	Parameter			
	$\overline{p}_i$ (MW)	x_i (h)	$\underline{p}_i$ (MW)	η_i (%)
1	3	4	4	70
2	3	2	3	60
3	6	1	3	90

discrete devices, each of the three DLR curves are piecewise linear. As a result, we envision that these metrics are passed between entities as vectors of their vertices; the lengths of these will scale as $\mathcal{O}(n)$.

Guaranteed recovery also means that multiple DLR packets can be sold ahead of operation for sequential use. The macro dispatcher need not fully utilize each DLR packet; in fact, it need not call upon it at all if the realized net demand does not exceed supply. Two options exist for a partial utilization of a purchased DLR packet: 1) between discharging and recharging operation, update E^r (retain y^* even though it may lead to an unnecessarily conservative rate constraint); 2) before the start of the operational period, allow for scheduling updates to reduce E^d and accordingly x^*.

V. NUMERICAL DEMONSTRATION

A. Example Construction of the DLR Curves

We now demonstrate construction of the DLR curves in a numerical example. Consider the fleet with parameters given in Table I, so that $n = 3$ and the other parameters are such that the maximum number of linear sections is observed on each of the discharge, loss and recovery time curves: 3, 3 and 5 respectively. Firstly, consider forming the DLR curves for just Battery 1. This can be seen in Figure 3 (upper panel). Next, consider aggregating the DLR capabilities of Batteries 1 and 2. This can be seen in Figure 3 (center panel). Finally, consider aggregating this compound DLR packet with that of Battery 3. The corresponding curves can be seen in Figure 3 (lower panel) and Figure 4.

B. Homogeneity Conditions

We here discuss which parameter values must be uniform across the fleet for it to be considered to be homogeneous. A homogeneous fleet has discharging *and* recovery curves that exactly encode the set of admissible signals and correspond to a single virtual battery.

1) Discharging Capacity Curve: The time-to-go x_i dictates the slope of the capacity curve over the corresponding section. To return a curve that is a straight line therefore requires equal x_i value across the fleet; this results in the discharging capabilities of a single virtual storage unit of power rating $\sum_{i=1}^n \overline{p}_i$ and total energy $\sum_{i=1}^n \overline{p}_i x_i$.

The DLR construction returns a (conditioned on E^d) discharging capacity curve that is a straight line between power rating $\sum_{i=1}^n \overline{p}_i$ and total energy $\sum_{i=1}^n \overline{p}_i \tilde{x}_i$.

2) Recharging Capacity Curve: Analogously to discharging mode, the requirement for a recharging capacity that corresponds to a single virtual battery is uniform time-to-charge y_i
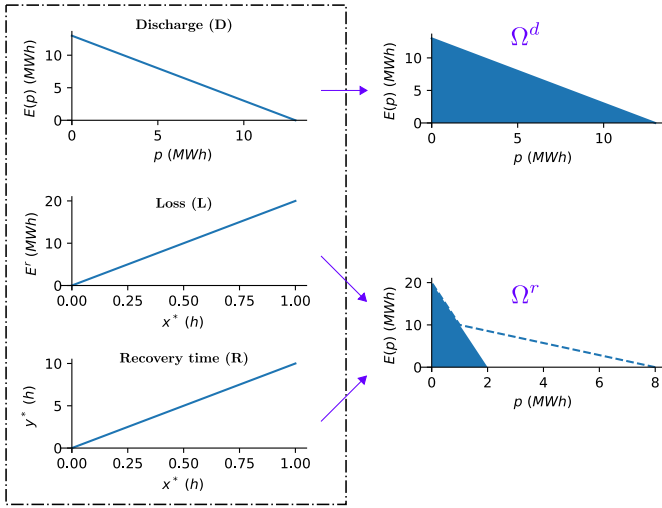
Fig. 6. A two-battery example in which the fleet homogeneity conditions are met, apart from the requirement for a uniform $\underline{p}_i$ value. The DLR curves are shown on the left and used to construct discharging and recharging capacity curves on the right for $E^d = \sum_{i=1}^{n} \bar{p}_i x_i$. Areas below the discharging and recharging capacity curves, as returned by the DLR formulation, are shaded in blue. For comparison, the full recharging capacity curve constructed from the fleet parameters is shown in dashed blue.

TABLE II
PARAMETERS FOR THE TWO-BATTERY EXAMPLE

Battery #	Parameter			
	$\bar{p}_i$ (MW)	x_i (h)	$\underline{p}_i$ (MW)	η_i (%)
1	7	1	7	70
2	6	1	1	60

across the fleet. When considering recharging between states of 0 and x_i (i.e., recovery from empty to full), this condition is equivalent to uniform $(\bar{p}_i x_i)/(\eta_i \underline{p}_i)$. The homogeneous recharge capacity curve has power rating $\sum_{i=1}^{n} \underline{p}_i$ and total energy $\sum_{i=1}^{n} \bar{p}_i x_i / \eta_i$.

When homogeneity of both discharging and recharging is enforced, the DLR construction returns a (conditioned on E^d) recharging capacity curve that is a straight line between power rating $\sum_{i=1}^{n} \underline{p}_i$ and total energy $\sum_{i=1}^{n} \bar{p}_i \tilde{x}_i / \eta_i$. This holds for any amount of energy dispatched.

As is to be expected, under these homogeneity conditions the loss and recovery time curves each take the form of a straight line. However, it should be noted that straight loss and recovery time curves are not sufficient conditions for homogeneity. By way of an example, Figure 6 shows the DLR curves for a 2-battery fleet with parameters given in Table II. This is not a homogeneous fleet, hence the recharging capacity curve formed via the DLR process is a truncation of the full-fleet curve. Note, however, the straight loss curve (due to uniform $\bar{p}_i x_i / \eta_i$) and recovery time curve (due to the same condition in combination with the $\max(\cdot)$ operator).

We use this example as an opportunity to point out the priority order that has been adopted in the definition of the DLR framework. We optimize for discharging flexibility, which takes precedence over recharging efficiency. We then prioritize simplicity in encoded recharging flexibility (fully described by

2 parameters) over recharging flexibility maximization, while guaranteeing minimum recharge time.

VI. BENCHMARKING THE DLR FRAMEWORK

We here qualitatively benchmark the DLR framework against the state-of-the art in optimal scheduling-plus-operation of aggregated storage fleets. The candidate approaches are described in the following subsections, and comparisons made in Table III. Note that, within each broad approach, we only consider options with feasibility guarantees for practical usability. In addition to the benefits shown in Table III, the DLR representation is a powerful tool for simplification. This will be shown in future work due to space constraints.

1) DLR Framework: In the scheduling phase, encode the DLR curves into an optimization problem to determine the reserved total energy E^d. During the operation phase, choose the aggregate power allocation in real time, subject to the equality constraint on total energy and *E-p* inequality constraints (scalar during recovery), and dispatch the response power over each sample period using the DLR broadcast procedure.

2) Sequential E-p Dispatch: During initial scheduling, reserve an *E-p* curve for the fleet (optionally: parameterize this via E^d and choose the maximum-flexibility truncation). Then, during discharge operation, allocate the power response over each sample period, subject to the *E-p* inequality constraints only, via broadcast dispatch. After discharging operation, transform into recharging coordinates and repeat the process.

3) Single Virtual Battery [4]–[7][4]: Choose a fixed set of scalar battery parameters to approximate the fleet, then incorporate this into the scheduling optimization problem. During operation, allocate a power level for each sample period subject to the scalar constraints on power and energy.

We point out here that it is straightforward to create a feasibility-guaranteeing virtual battery from the DLR curves for a given choice of E^d.

4) Asymmetric Block Offers [17][4]: Compute a finite set of discharge-recharge offers, each approximating the fleet as a different single virtual battery. Encode all of these offers into an optimization problem covering the full operational period. During operation, allocate a power level for each sample period subject to the scalar constraints on power and energy.

5) Full Information Scheduling [2], [3]: Ahead of time, use full information for each device to determine an allocation of the fleet as the solution to an optimization problem covering the full operational period. Dispatch the solution ahead of time.

6) Full Information Direct Control [13], [14]: Omit the scheduling phase. In real time, communicate directly with each device to achieve the optimal response.

VII. APPLICATION CONSIDERATIONS

A. Demand Response

While the DLR framework is motivated by the capabilities of a fleet of storage devices, it can be naturally applied to demand

---

[4]We consider an inner bound on parameters such that the single virtual battery encoding is guaranteed to result in the allocation of feasible signals

TABLE III
HIGH-LEVEL DESCRIPTION: BENCHMARKING THE DLR FRAMEWORK AGAINST ALTERNATIVE DISPATCH APPROACHES.
LOW-LEVEL DESCRIPTION: GREEN CELLS ARE PREFERRED OVER ORANGE, WHICH ARE IN TURN PREFERRED OVER RED.
WITHIN THE GREEN CELLS, DARK GREEN DENOTE AN IMPROVEMENT OVER LIGHT GREEN

Approach	discharging flexibility	advance commitment	energy recovery	recharging flexibility	aggregation of products	visual representation of flexibility	dispatch method
DLR framework	full*	discharge energy	exact, in advance	minimum time	yes	yes	broadcast
Sequential E-p dispatch	full*	none	exact, after discharge	full*	yes	yes	broadcast
Single virtual battery**	limited, continuous	none	not guaranteed	limited, continuous	no	yes	implementation dependent
Asymmetric block offers**	limited, options given	full	exact, in advance	limited, options given	no	yes	implementation dependent
Full information scheduling	full	full	exact, in advance	full	yes	no	direct, ahead of time
Full information direct control	full	none	exact, full control	full	yes	no	direct, real time

* with the exclusion of cross-charging as discussed in Section II.
** parameters chosen to ensure feasibility in both directions.

TABLE IV
MODELING A RANGE OF DEMAND RESPONSE CHARACTERISTICS
VIA THE PARAMETER η_i

Load characteristic	η_i value
Pure demand shifting	$\eta_i = 1$
Demand shifting with excess recovery	$0 < \eta_i < 1$
Demand shifting with incomplete recovery	$1 < \eta_i < \infty$
Pure demand turn-down (no recovery)	$\eta_i = \infty$

response by considering each flexible load as a *virtual storage unit*. In particular, this is then well suited to the aggregation of portfolios of heterogeneous flexible loads. An additional benefit here is that the framework can be used to model a variety of load characteristics (demand turn-down, demand shifting with specific recovery requirements), via the chosen value of η_i, as can be seen in Table IV; at which point the heterogeneous combination can be aggregated as described. Note that, in order to model demand destruction, we extend the domain of η_i beyond the limits of our physical battery model $\eta_i \in (0, 1]$. The framework includes the encoding of minimum discharge (x^*) and recharge (y^*) times, but in order to ensure timely recovery of shiftable demand, a maximum allowable completion time may need to be communicated in addition.

B. System Balancing

The DLR framework is well suited to system balancing, in particular to markets where there are two distinct phases: reservation followed by dispatch. An example of such a market is the aFRR (automatic frequency restoration reserve) product that is common across European TSOs under the ENTSO-E umbrella organization [27].

We focus our attention on the implementation of aFRR by TenneT TSO (Netherlands) [41]. It utilize a mixture of pre-purchased contracts (to ensure the ability to meet minimum requirements) and just-in-time market clearing to *reserve* response, ahead of each 15 min operational period. Bids are submitted via the BRP, up to 25 min before delivery, and are separated into upwards and downwards regulation. The TSO creates a merit order list by gate closure and submits this to the aFRR platform. Instructions are then issued to *dispatch* reserved flexibility within each 15 min operational period, with a response time of 30 s. The aFRR platform sends dispatch instructions to TSO-local aFRR controllers, every 1-10 s. This service is not strictly limited to large assets; in fact a number of European TSOs are developing a platform that aims to
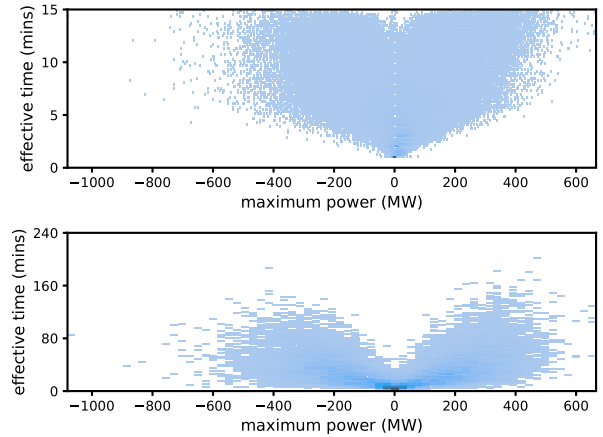

Fig. 7. The distribution of energy (normalized to time) as a function of maximum power, across the Tennet dispatch dataset, for request windows of 15 min (upper panel) and 4 h (lower panel). Up-regulation requests are shown as positive and down-regulation requests as negative.

enable aggregation of small-scale resources to deliver aFRR services (Equigy, launched as a joint venture between TenneT, SwissGrid and Terna [28]). It is conceivable that such a service may in the future also deliver local congestion management.

In order to assess the suitability of the DLR framework for application to the TenneT Netherlands aFRR market, we analyze the reported dispatch data for the 5 y period covering 2017 through 2021, at 1 min granularity, publicly available from the TenneT website [42]. Currently, the market requires block bids of constant power per PTU (15 min). We will investigate the possible benefits of more flexible products for such a service. To do so, we split the 5 y traces into request segments, at 15 min as well as 4 h granularity to assess longer-duration requirements (e.g., for pre-contracted services). We analyze up- and down-regulation separately, because these are treated as independent products; for each we neglect periods that have the value 0 MW throughout their duration. Firstly, we investigate the distribution of dispatched energy versus the maximum power level within each time window. The results of this can be seen in Figure 7, where the vertical axis shows the dispatched energy divided by maximum power – a duration that is necessarily capped at the window size. Figure 8 shows the projection of these densities on the effective-time axis, separately for up- and down-regulation.

As can be seen in these figures, it is very rare for the observed peak power to be required across the entire settlement period; rather, a large range of effective time spans is observed. This is true across all peak power levels and the
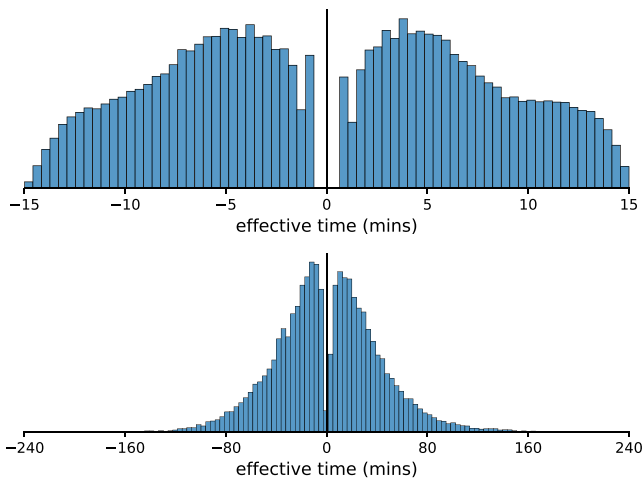
Fig. 8. The distribution of effective request times, for request windows of 15 min (upper panel) and 4 h (lower panel). Times corresponding to up-regulation requests are shown as positive and down-regulation as negative.

TABLE V
HIGH-LEVEL DESCRIPTION: CORRELATION COEFFICIENTS
BETWEEN MAXIMUM POWER AND EFFECTIVE TIME
FOR THE RESULTS OF FIGURE 7.
LOW-LEVEL DESCRIPTION: NEGATIVE VALUES RESULT FROM THE SIGN
CONVENTION THAT DOWN-REGULATION REQUESTS ARE NEGATIVE

Window size	Service direction	
	Up-regulation	Down-regulation
15 min	0.64	-0.59
4 h	0.62	-0.59

effective time tends to increase with maximum power magnitude (correlation coefficients are shown in Table V). While for 15 min observation windows the effective time can take values up to the full duration (i.e., a constant power requirement), the 4 h requests do not exceed an effective time of 203 min and the majority have an energy requirement corresponding to less than half of the period length.

These results suggests that, under the currently used framework, the TSO is required to purchase a wide variety of different power-energy products, given the restriction to block offers in which the energy is given by 15 min multiplied by the fixed power. We propose an improved alternative to this: that it instead purchases a single (compound) DLR product that is a better match for its needs for the period. To demonstrate the potential benefits of this approach over the restriction to block offers of fixed duration, we plot the discharge capacity curve across 50 randomly-sampled dispatch signals from the TenneT dataset in Figure 9. We do this in both the up- and down-regulation directions, and for each sample normalize the request so that its maximum power is equal to 1; in this way we can visually compare across different request magnitudes.

As it can be seen in Figure 9, the discharge capacity curves lie below the straight line corresponding to a single block offer, covering more of this space in the 15 min case than in the 4 h case. By construction, the distribution in Figure 8 corresponds to the distribution of the y-axis intercepts of these (rescaled) *E-p* transforms. The combination of these results shows that, for 15 min request windows, the majority of requests use less than the full energy that is reserved in the block offer. For
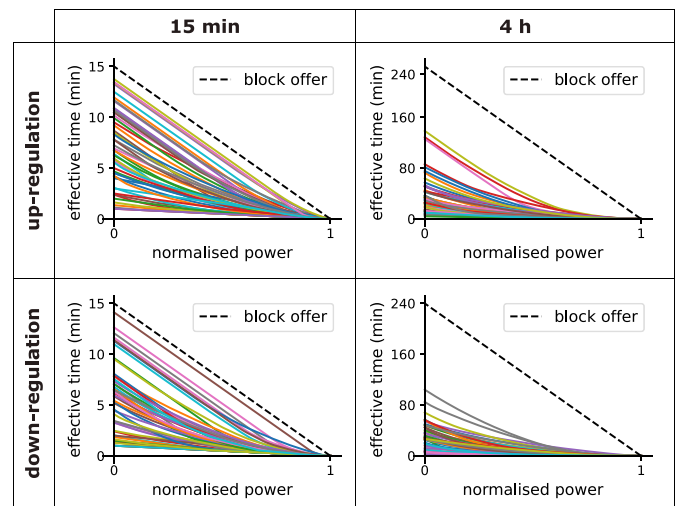

Fig. 9. Discharge capacity curves composed from the Tennet dispatch request dataset. 50 random samples are shown for each of up-regulation at 15 min granularity (upper left panel), down-regulation at 15 min granularity (lower left panel), up-regulation at 4 h granularity (upper right panel) and down-regulation at 4 h granularity (lower right panel).

4 hour request windows, the mismatch between power and energy is exacerbated. This suggests that advance balancing contracts could make allowances for energy constraints and, for example, be provided by duration-limited batteries.

To summarize, use of the discharge capacity curve in place of (a stack of diverse) block offers allows the system operator to procure the exact reserve that it requires, in one well-defined and transparent interaction with aggregators. Use of the DLR approach captures this benefit and in addition extends it into settings of response followed by recovery. The utilization of energy-limited assets (physical batteries or virtual batteries formed of, for example, flexible loads) for balancing necessitates such a recovery mode. This is challenging for system balancing because activation will either prohibit rapid recovery or cause uncontrolled recovery (which causes further balancing concerns). The DLR approach lets the balancing authority explicitly control round-trip operation and dispatch the energy recovery process. This may result in additional efficiency by reducing the need to purchase regulation products in the opposing direction.

C. Dealing With Uncertainties

As presented thus far, the DLR framework is a deterministic methodology. A DLR packet is computed to encode the fleet capabilities and compared to system needs, for example via embedding into an optimization problem. The same framework can also be applied in stochastic settings, by taking a scenario approach and performing this comparison for each scenario. This leads to two possible application settings: 1) a range of DLR packets is computed, sampling uncertainty in the capabilities of the fleet, and a determination made on the ensemble of DLR curves (stochastic capabilities); 2) a single DLR packet is computed and compared to an ensemble of different system requirements (stochastic system needs). A combination of the two would also be possible, in which both the capabilities and system needs were sampled from distributions.

By way of an example, we demonstrated chance-constrained optimization under stochastic capabilities for unidirectional operation, using the capacity curve alone, in [23], via a quantile approach. The same approach could be used to make determinations on stochastic fleet capabilities for response-recovery operation using the DLR curves. As the constituent curves (previously the discharging curve alone) serve to encode the flexibility constraints in any given scenario, this approach leads to a consistent deduction of the stochastic flexibility bounds across the range of scenarios sampled.

VIII. Conclusion and Future Work

This paper has presented a framework for the procurement of reserve flexibility with guaranteed recovery. The DLR representation encodes full discharging capabilities and easily interpretable recovery capabilities that exactly account for losses. Feasibility is guaranteed, so that any request deemed admissible based on the purchased reserve can necessarily be met by the fleet. We have provided an example construction of the constituent curves and benchmarked our approach against the state-of-the-art.

In future work, the authors intend to extend these results into settings where cross-charging is allowed to take place. In particular, they aim to investigate the applicability of the DLR framework in such circumstances and characterize any resulting reduction in optimality. In addition, the possibility to use simplified fleet representations (which lead to even more compact DLR representations) will be investigated.

Appendix

A. Proof of Theorem 1

Given $\tilde{x}$, define $\tilde{p} \doteq \sum_{i:\, x_i \geq \tilde{x}} \overline{p}_i$. Using this, we can write down the capacity curve based on $\tilde{x}$ as follows:

$$\Omega_{\overline{p}, \tilde{x}}(p) = \begin{cases} \Omega_{\overline{p}, x}(p), & \text{if } p \geq \tilde{p} \\ \Omega_{\overline{p}, x}(\tilde{p}) + (\tilde{p} - p)\tilde{x}, & \text{otherwise.} \end{cases} \quad (A.1)$$

In particular, $\Omega_{\overline{p}, \tilde{x}}(0) = E^d$. We can therefore deduce that

$$\left. \begin{array}{l} E_{pd}(p) \leq \Omega_{\overline{p}, x}(p) \quad \forall p \\ E_{pd}(0) \leq E^d \end{array} \right\} \iff E_{pd}(p) \leq \Omega_{\overline{p}, \tilde{x}}(p) \; \forall p, \quad (A.2)$$

and so the result follows from Property 1.

B. Proof of Theorem 2

Consider two fleets of devices: $(\overline{p}^{(1)}, x^{(1)})$ of size n, indexed by descending time-to-go values (i.e., $x_i \geq x_{i+1}$) and $(\overline{p}^{(2)}, x^{(2)})$ that consists of a single device. With slight abuse of notation, we denote the parameters of the latter by $\overline{p}^{(2)}, x^{(2)}$. We define j: $x_j^{(1)} \geq x^{(2)} > x_{j+1}^{(1)}$, with $x_0^{(1)} \doteq x^{(2)}$ and $x_{n+1}^{(1)} \doteq 0$. Now, consider the resultant capacity curve of the combined fleet, as follows:

$$\Omega_{1+2}(p) = \begin{cases} \Omega_1(p) + \overline{p}^{(2)} x^{(2)}, & \text{if } p \leq p' \\ \Omega_1(p - \overline{p}^{(2)}), & \text{if } p \geq p' + \overline{p}^{(2)} \\ \Omega_1(p') + (p' + \overline{p}^{(2)} - p)x^{(2)}, & \text{otherwise,} \end{cases}$$

$$(A.3)$$

in which $p' \doteq \sum_{i=1}^{j} \overline{p}_i^{(1)}$. Denote the flexibilities corresponding to the two fleets as $\mathcal{F}_1$ and $\mathcal{F}_2$. Denote their vertices as the sets $\mathcal{V}_1$ and $\mathcal{V}_2$, which are given by

$$\mathcal{V}_1 = \mathcal{C}(\Omega_1) \cup \{(0, +\infty), (+\infty, 0)\}, \quad (A.4a)$$
$$\mathcal{V}_2 = \left\{ (0, +\infty), \left(0, \overline{p}^{(2)} x^{(2)}\right), \left(\overline{p}^{(2)}, 0\right), (+\infty, 0) \right\}, \quad (A.4b)$$

in which $\mathcal{C}(\Omega)$ denotes the vertices of a capacity curve $\Omega(\cdot)$. We can then compute the Minkowski sum as follows:

$$\begin{aligned} \mathcal{F}_1^{\mathsf{C}} \oplus \mathcal{F}_2^{\mathsf{C}} &= \text{Conv}(\{a + b : a \in \mathcal{V}_1, \; b \in \mathcal{V}_2\}) \\ &= \text{Conv}\Big(\Big\{ a + b : a \in \mathcal{C}(\Omega_1), \; b = \Big(0, \overline{p}^{(2)} x^{(2)}\Big), \Big(\overline{p}^{(2)}, 0\Big) \Big\} \\ &\qquad \cup \{(0, +\infty), (+\infty, 0)\} \Big), \quad (A.5) \end{aligned}$$

in which $\text{Conv}(\cdot)$ denotes the formation of the convex hull. This returns the corresponding set of vertices

$$\begin{aligned} & \left\{ \left(\sum_{i=1}^{k} \overline{p}^{(1)}, \sum_{i=k}^{n} \overline{p}_k^{(1)} x_k + \overline{p}^{(2)} x^{(2)} \right), \; k = 0, \ldots, j-1 \right\} \\ & \cup \left\{ \left(\sum_{i=1}^{k} \overline{p}^{(1)} + \overline{p}^{(2)}, \sum_{i=k}^{n} \overline{p}_k^{(1)} x_k \right), \; k = j, \ldots, n \right\} \\ & \cup \{(0, +\infty), (+\infty, 0)\} \\ & = \mathcal{C}(\Omega_{1+2}) \cup \{(0, +\infty), (+\infty, 0)\}, \quad (A.6) \end{aligned}$$

so the result follows for the addition of a single device to a complex fleet. The general result then follows because the Minkowski summation operator is associative.

References

[1] "Future energy scenarios," Nat. Grid, London, U.K., Rep. FES 2021, Jul. 2021. [Online]. Available: https://www.nationalgrideso.com/document/202851/download

[2] A. Y. Saber and G. K. Venayagamoorthy, "Resource scheduling under uncertainty in a smart grid with renewables and plug-in vehicles," *IEEE Syst. J.*, vol. 6, no. 1, pp. 103–109, Mar. 2012.

[3] N. G. Paterakis, O. Erdinç, I. N. Pappi, A. G. Bakirtzis, and J. P. S. Catalao, "Coordinated operation of a neighborhood of smart households comprising electric vehicles, energy storage and distributed generation," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2736–2747, Nov. 2016.

[4] S. Vandael, B. Claessens, M. Hommelberg, T. Holvoet, and G. Deconinck, "A scalable three-step approach for demand side management of plug-in hybrid vehicles," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 720–728, Jun. 2013.

[5] M. G. Vayá and G. Andersson, "Optimal bidding strategy of a plug-in electric vehicle aggregator in day-ahead electricity markets under uncertainty," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2375–2385, Sep. 2015.

[6] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "Evaluation of achievable vehicle-to-grid capacity using aggregate PEV model," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 784–794, Jan. 2017.

[7] Z. Xu, D. S. Callaway, Z. Hu, and Y. Song, "Hierarchical coordination of heterogeneous flexible loads," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4206–4216, Nov. 2016.

[8] A. Mehrabi, H. S. V. S. K. Nunna, A. Dadlani, S. Moon, and K. Kim, "Decentralized greedy-based algorithm for smart energy management in plug-in electric vehicle energy distribution systems," *IEEE Access*, vol. 8, pp. 75666–75681, 2020.

[9] M. Mazidi, A. Zakariazadeh, S. Jadid, and P. Siano, "Integrated scheduling of renewable generation and demand response programs in a microgrid," *Energy Conv. Manag.*, vol. 86, pp. 1118–1127, Oct. 2014.

[10] X. Wang, A. Palazoglu, and N. H. El-Farra, "Operational optimization and demand response of hybrid renewable energy systems," *Appl. Energy*, vol. 143, pp. 324–335, Apr. 2015.

[11] K. M. Tsui and S. C. Chan, "Demand response optimization for smart home scheduling under real-time pricing," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1812–1821, Dec. 2012.

[12] Z. Wang, C. Gu, F. Li, P. Bale, and H. Sun, "Active demand response using shared energy storage for household energy management," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1888–1897, Dec. 2013.

[13] F. Adamek, M. Arnold, and G. Andersson, "On decisive storage parameters for minimizing energy supply costs in multicarrier energy systems," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 102–109, Jan. 2014.

[14] T. Pippia, J. Sijs, and B. De Schutter, "A single-level rule-based model predictive control approach for energy management of grid-connected microgrids," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 6, pp. 2364–2376, Nov. 2020.

[15] C. A. Correa-Florez, A. Michiorri, and G. Kariniotakis, "Optimal participation of residential aggregators in energy and local flexibility markets," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1644–1656, Mar. 2020.

[16] J. T. Hughes, A. D. Domínguez-García, and K. Poolla, "Identification of virtual battery models for flexible loads," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4660–4669, Nov. 2016.

[17] N. O'connell, P. Pinson, H. Madsen, and M. O'malley, "Economic dispatch of demand response balancing through asymmetric block offers," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2999–3007, Jul. 2016.

[18] M. Amini and M. Almassalkhi, "Optimal corrective dispatch of uncertain virtual energy storage systems," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4155–4166, Sep. 2020.

[19] H. T. Nguyen, A. S. Al-Sumaiti, K. Turitsyn, Q. Li, and M. S. El Moursi, "Further optimized scheduling of micro grids via dispatching virtual electricity storage offered by deferrable power-driven demands," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3494–3505, Sep. 2020.

[20] R. R. Appino, V. Hagenmeyer, and T. Faulwasser, "Towards optimality preserving aggregation for scheduling distributed energy resources," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 3, pp. 1477–1488, Sep. 2021.

[21] M. P. Evans, S. H. Tindemans, and D. Angeli, "A graphical measure of aggregate flexibility for energy-constrained distributed resources," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 106–117, Jan. 2020.

[22] J. Cruise and S. Zachary, "Optimal scheduling of energy storage resources," 2018, *arXiv:1808.05901*.

[23] M. P. Evans, D. Angeli, G. Strbac, and S. H. Tindemans, "Chance-constrained ancillary service specification for heterogeneous storage devices," in *Proc. IEEE PES Innov. Smart Grid Technol. Eur. (ISGT-Europe)*, 2019, pp. 1–5.

[24] "Ancillary Services." PJM. 2017. [Online]. Available: http://www.pjm.com/markets-and-operations/ancillary-services.aspx

[25] "Load Participation and Demand Response." CAISO. 2017. [Online]. Available: http://www.caiso.com/participate/Pages/Load/Default.aspx

[26] "Balancing Services." National Grid. 2020. [Online]. Available: https://www.nationalgrideso.com/balancing-services

[27] "ENTSO-E automatic frequency restoration reserve process," ENTSO-E, Brussels, Belgium, Rep., 2019. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/EDI/Library/ERRP/Automatic_Frequency_Restoration_Reserve_Process_IG_v1.0.pdf

[28] "Equigy Platform Gives European Consumers Access to Tomorrow's Sustainable Energy Market." TenneT. 2020. [Online]. Available: https://www.tennet.eu/tinyurl-storage/detail/equigy-platform-gives-european-consumers-access-to-tomorrows-sustainable-energy-market/

[29] V. Trovato, S. H. Tindemans, and G. Strbac, "Leaky storage model for optimal multi-service allocation of thermostatic loads," *IET Gener. Transm. Distrib.*, vol. 10, no. 3, pp. 585–593, 2016.

[30] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Aggregate flexibility of thermostatically controlled loads," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 189–198, Jan. 2015.

[31] L. Zhao, W. Zhang, H. Hao, and K. Kalsi, "A geometric approach to aggregate flexibility modeling of thermostatically controlled loads," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4721–4731, Nov. 2017.

[32] Z. Yi, Y. Xu, W. Gu, L. Yang, and H. Sun, "Aggregate operation model for numerous small-capacity distributed energy resources considering uncertainty," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4208–4224, Sep. 2021.

[33] S. Kundu, K. Kalsi, and S. Backhaus, "Approximating flexibility in distributed energy resources: A geometric approach," in *Proc. 20th Power Syst. Comput. Conf. (PSCC)*, 2018, pp. 1–7.

[34] F. L. Müller, J. Szabó, O. Sundström, and J. Lygeros, "Aggregation and disaggregation of energetic flexibility from distributed energy resources," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1205–1214, Mar. 2019.

[35] A. Ulbig and G. Andersson, "Analyzing operational flexibility of electric power systems," *Int. J. Electr. Power Energy Syst.*, vol. 72, pp. 155–164, Nov. 2015.

[36] T. Li, B. Sun, Y. Chen, Z. Ye, S. H. Low, and A. Wierman, "Learning-based predictive control via real-time aggregate flexibility," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 4897–4913, Nov. 2021.

[37] G. Tsaousoglou, K. Mitropoulou, K. Steriotis, N. G. Paterakis, P. Pinson, and E. Varvarigos, "Managing distributed flexibility under uncertainty by combining deep learning with duality," *IEEE Trans. Sustain. Energy*, vol. 12, no. 4, pp. 2195–2204, Oct. 2021.

[38] S. Zachary, S. H. Tindemans, M. P. Evans, J. R. Cruise, and D. Angeli, "Scheduling of energy storage," *Philos. Trans. Roy. Soc. A, Math. Phys. Eng. Sci.*, vol. 379, no. 2202, 2021, Art. no. 20190435.

[39] M. Evans, S. H. Tindemans, and D. Angeli, "Robustly maximal utilisation of energy-constrained distributed resources," in *Proc. Power Syst. Comput. Conf. (PSCC)*, 2018, pp. 1–7.

[40] M. P. Evans, S. H. Tindemans, and D. Angeli, "Minimizing unserved energy using heterogeneous storage units," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3647–3656, Sep. 2019.

[41] "Product information aFRR (regulating power)," TenneT, Arnhem, The Netherlands, Rep., 2018. [Online]. Available: https://www.tennet.eu/fileadmin/user_upload/Company/Publications/Technical_Publications/Dutch/Product_information_aFRR_-regulating_power-_16-01-2018.pdf

[42] "Export Data." TenneT. 2022. [Online]. Available: https://www.tennet.org/english/operational_management/export_data.aspx

Michael P. Evans received the B.A. and M.Eng. degrees in engineering from the University of Cambridge, U.K., in 2015, and the Ph.D. degree in electrical and electronic engineering from Imperial College London, U.K., in 2019.

He is currently with Octopus Energy, London, U.K. His research interests lie in the optimal control of power systems, with a particular focus on the provision of aggregate flexibility from large fleets of heterogeneous assets.

Simon H. Tindemans (Member, IEEE) received the M.Sc. degree in physics from the University of Amsterdam, Amsterdam, The Netherlands, in 2004, and the Ph.D. degree in theoretical biophysics from Wageningen University, Wageningen, The Netherlands, in 2009.

From 2010 to 2017, he was a Research Associate, a Marie-Curie Intra-European Fellow, and a Research Fellow with the Control and Power Group, Imperial College London, U.K. He is currently an Assistant Professor with the Department of Electrical Sustainable Energy, Delft University of Technology, Delft, The Netherlands. His research concerns the use of statistical approaches to model, analyze, and control the power system. Within this domain, he has a specific research interest in the aggregate behavior of large numbers of heterogeneous and unpredictable resources, and the ability to characterize and control their aggregate flexibility.

David Angeli (Fellow, IEEE) received the B.S. degree in computer science engineering and the Ph.D. degree in control theory from the University of Florence, Italy, in 1996 and 2000, respectively.

Since 2000, he has been an Assistant and Associate Professor with the Department of Information Engineering, University of Florence in 2005. In 2008, he joined the EEE Department, Imperial College London, where he is currently a Professor of Nonlinear Network Dynamics and the Director of the M.Sc. in Control Systems. He has authored more than 100 journal papers in the areas of Stability of nonlinear systems, Control of constrained systems (MPC), Chemical Reaction Networks theory, and Smart Grids.

Prof. Angeli was a recipient of the 2021 International Honeywell Medal from the Institute of Measurement and Control. He has been a Fellow of IET since 2018 and a Fellow of IEEE since 2015