2012 American Control Conference
Fairmont Queen Elizabeth, Montréal, Canada
June 27-June 29, 2012

Real-time Scheduling of Deferrable Electric Loads

A. Subramanian*‡, M. Garcia*, A. Domínguez-García¶, D. Callaway§*, K. Poolla*†, P. Varaiya†

***Abstract*— We consider a collection of distributed energy resources [DERs] such as electric vehicles and thermostatically controlled loads. These resources are flexible: they require delivery of a certain total energy over a specified service interval. This flexibility can facilitate the integration of renewable generation by absorbing variability, and reducing the reserve capacity and reserve energy requirements. We first model the energy needs of these resources as tasks, parameterized by arrival time, departure time, energy requirement, and maximum allowable servicing power. We consider the problem of servicing these resources by allocating available power using real-time scheduling policies. The available generation consists of a mix of renewable energy [from utility-scale wind-farms or distributed rooftop photovoltaics], and load-following reserves. Reserve capacity is purchased in advance, but reserve energy use must be scheduled in real-time to meet the energy requirements of the resources. We show that there does not exist a causal optimal scheduling policy that respects servicing power constraints. We then present three heuristic causal scheduling policies: Earliest Deadline First [EDF], Least Laxity First [LLF], and Receding Horizon Control [RHC]. We show that EDF is optimal in the absence of power constraints. We explore, via simulation studies, the performance of these three scheduling policies in the metrics of required reserve energy and reserve capacity.**

I. INTRODUCTION

The worldwide interest in renewable energy is driven by pressing environmental problems, energy supply security issues, and nuclear power safety concerns. However, the widespread adoption of renewables presents serious operational challenges. Renewable energy sources such as wind and solar are fundamentally different from conventional generation such as coal, nuclear, and natural gas. The energy production from these renewable sources is not dispatchable [cannot be controlled on demand], is intermittent [exhibits large fluctuations], and is uncertain [random or not known in advance]. We use the phrase *variable generation* [VG] to encompass these three characteristics of renewable energy sources [20]. Variability is the most important obstacle to deep integration of renewable generation into the electric energy system.

The current approach to renewable integration is to absorb the attendant variability in operating reserves. This works at today's modest penetration levels, but it will not scale tomorrow, when we have 30% or more of the energy consumption drawn from renewables. Recent studies in California [8], [15] project that the spring time maximum up-regulation capacity needed to accommodate 33% renewable energy penetration will increase from 277 MW to 1,135 MW. Similar increases are projected in required down-regulation capacity. Maximum load-following capacity requirements will need to increase from 2,292 MW to 4,423 MW. These large increases in necessary reserves are economically untenable, and defeat the net carbon benefit from deploying renewable energy sources.

Enabling deep penetration of variable generation will require the convergence of a number of solutions including establishing intraday markets [5] to leverage better forecasts in shorter horizons, curtailment of renewable generation to limit injected variability [6], and *coordinated resource aggregation*.

Coordinated Aggregation is substantially more powerful than traditional demand response [DR]. It (a) involves the intelligent control of resources such as deferrable loads and available storage to match variable generation, and (b) it creates bulk power and ancillary service market opportunities far more substantial than those offered by the peak-shaving capacity markets. Existing DR research has focused on the limited market opportunities of (a) shifting load from peak [17], [18], (b) contingency reserves [14], [25], and (c) decentralized frequency response [13], [24].

The value of coordinated aggregation rests on the improved forecastability it offers in *ex-ante* markets. This, in turn, reduces capacity reserve requirements needed for load following. There are two approaches to coordination that have been studied in the literature: (a) direct-load control [16], where deferrable loads are centrally managed, and (b) indirect load-control [10], [11], [19], where deferrable loads respond to generation conditions through the proxy of real-time prices. We submit that direct-load control is the more attractive option as it offers greater flexibility and would face fewer adoption obstacles as customers are not exposed to real-time price volatility. Indeed, there is, to our knowledge, no current implementation where residential customers are offered and respond to real-time pricing.

Our central construct is that of a *resource cluster*. This is a diverse collection of networked resources at the distribution side including renewable and micro-generation, deferrable loads, and electricity storage. Resource clusters are managed by a *Cluster Manager* [CM] which aggregates the cluster's capabilities and presents them to the system operator [SO] as a dispatchable resource. A CM-based approach to the control of deferrable loads is necessary because centralized control of resources falls outside the purview and business models of

Supported in part by the UC Discovery Grant ele07-10283 under the IMPACT program, EPRI and CERTS under sub-award 09-206, NSF under Grants EECS-0925337 and 1129001, and Robert Bosch LLC through its Bosch Energy Research Network funding program.

‡ Corresponding author: anandsub@berkeley.edu
* Mechanical Engineering, U.C. Berkeley
† Electrical Engineering and Computer Science, U.C. Berkeley
§ Energy and Resources Group, U.C. Berkeley
¶ Electrical and Computer Engineering, University of Illinois, Urbana

system operators, and because the associated computational costs are prohibitive [9].

The CM has two functions: (a) *ex-ante* acquisition of power from various generation sources. These include grid power, distributed renewables [ex: rooftop PVs], and utility-scale renewables [ex: wind farms] in traditional electricity markets. The amount of power acquired is based on load forecasts. Additionally, reserve power capacity is purchased based on load and generation forecast statistics. (b) Real-time scheduling of power to various deferrable loads. Each load relays its energy needs to the CM. The CM uses all available information to compute power schedules which are then communicated to the loads. Loads then configure their hardware to accept allocated power levels. Shortfalls in required energy to service these loads are acquired from reserve generation. However, ex-ante reserve capacity decisions constrain the amount of required energy that can be accepted from reserve generation. Clearly, this vision requires Smart Grid communication and computation infrastructure to enable efficient data transfer between resources and the CM. This paper focuses on the real-time scheduling function of the CM.

The remainder of this paper is organized as follows. Section II contains the problem formulation where we model flexible loads, describe generation acquisition, and introduce scheduling policies. In Section III, we show that there does not exist a causal optimal scheduling policy under rate constraints, and describe three different causal heuristic policies: Earliest Deadline First [EDF], Least Laxity First [LLF], and Receding Horizon Control [RHC]. In Section IV, we compare the performance of these scheduling policies to the baseline case of no resource coordination. In Section V, we draw conclusions and suggest future research directions.

II. Problem Set-up

A. Resource Modeling

Consider a collection of distributed energy resources: deferrable loads such as electric vehicles [EVs], and thermostatically controlled loads [TCLs] such as HVACs or refrigeration units. We do not consider electricity storage in this paper. Detailed models of such resources can be readily constructed to varying levels of fidelity [11]. The energy needs of each of these resources can be regarded [possibly conservatively] as *tasks*.

Definition 1 (Task): A task is characterized by a service interval $I = [a, d]$ over which a total energy amount E must be delivered with a maximum power transfer rate of m. Let $p(t)$ be the power consumed by a task. The task energy requirement is then:

$$\int_a^d p(t)dt = E, \quad 0 \leq p(t) \leq m. \tag{1}$$

Each task is parametrized by (E, a, d, m). For an EV, E is the energy required to charge its battery on $[a, d]$ to a user-specified level. For a TCL, E is the energy required to maintain temperature within a specified range over the duty-cycle $[a, d]$. This depends on exogenous variables such as ambient temperature and user comfort.

We assume the consumed power $p(t)$ can take on any value on the continuous interval $[0, m]$. This simplifies analysis and enables the development of efficient scheduling algorithms. For practical implementation of our algorithms, we propose a simple rounding procedure where $p(t)$ is set to the nearest available discretized power level.

We allow tasks to be *pre-emptive*, that is they can be serviced with interruptions. Scheduling problems for non-pre-emptive tasks reduce to bin packing problems which are NP hard [1].

B. Generation Modeling

The available power $p(t)$ to service various tasks is drawn from:

(a) Renewable generation $w(t)$: This refers to variable generation from both utility-scale wind farms and distributed rooftop PVs. We assume renewable generation is *free* because such sources of power have small marginal costs of production.

(b) Grid generation $g(t)$: This includes energy purchased in traditional forward markets [bulk power] and ancillary services markets [reserve energy]. Grid generation, in contrast to renewable generation, is *certain but costly*. The typical procurement timeline for each component of $g(t)$ is illustrated in Figure 1 below. Bulk power (B) and reserve capacity (C) are purchased in day-ahead, and hour-ahead forward electricity markets. Alternatively, reserve energy $(r(t))$ use must be decided in real-time to meet load energy requirements. There are a great many nuanced but essential details on the process of acquiring available power that we do not consider in this paper.

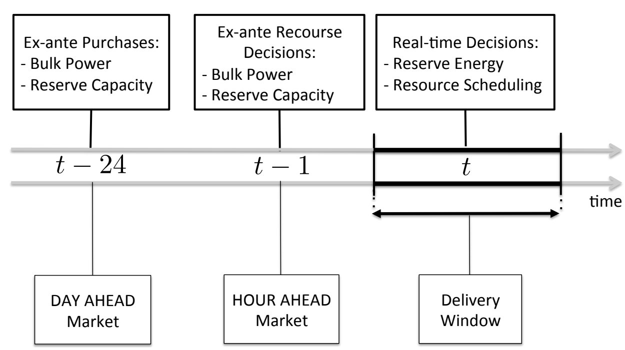


Fig. 1: Typical timeline for power acquisition

C. Cost Metric

We focus primarily on the real-time scheduling of tasks. The CM must fulfill all task requirements by acquiring adequate renewable and grid generation. Thus, the cost of acquired generation is the appropriate performance metric for comparing different task scheduling policies. In this paper, we assume there are two components to this cost:

$$\int_0^T |g(t)|dt + \alpha \max_t |g(t)|. \quad (2)$$

The first term in (2) penalizes the total energy procured from grid generation [grid energy] while the second captures the capacity cost of having grid generation capability [grid capacity]. The parameter α negotiates the relative importance of these costs.

D. Scheduling Policies

We now introduce some definitions which we will require in the remainder of the paper.

Definition 2 (Energy state, Active task): Consider a task T parametrized by (E, a, d, m). Suppose $p(t)$ is the power profile allocated to this task. The *energy state* of task T at time t is:

$$e(t) = E - \int_a^t p(\tau)d\tau.$$

The task T is called *active at time* t if $a \le t \le d$ and $e(t) > 0$. Consider a collection of tasks $\mathbb{T} = \{T_i\}_{i=1}^M$, where T_i is parameterized as (E_i, a_i, d_i, m_i). The set of active tasks at time t is written $\mathbb{A}_t$.

We allocate an *available power profile* $p(t)[= w(t) + g(t)]$ to service these tasks. We have access to the following *information state*.

Definition 3 (Information State): The information state $\mathcal{I}_t$ at time t consists of:

(a) Parameters (E_i, a_i, d_i, m_i) for all tasks T_i active at time t.

(b) Energy states $e_i(t)$ of all tasks T_i active at time t.

(c) Realized values of the available power profile, i.e. $p(\tau) : \tau \le t$.

Definition 4 (Scheduling Policy): A scheduling policy σ is an algorithm that allocates available power profile to active tasks. Specifically, for a collection $\mathbb{T}$ of M tasks, let $\sigma(p, t) = (p_1(t), \cdots, p_M(t))$, where $p_i(t)$ is the power allocated to task i at time t. Clearly, we require,

$$\sum_{i=1}^M p_i(t) \le p(t).$$

A policy σ is called *causal* if its allocation at time t depends only on the information state $\mathcal{I}_t$.

Definition 5 (Feasibility): The available power profile $g(t)$ is called *feasible* if there exists some [*possibly non-causal*] scheduling policy that completes the tasks:

$$e(d_i) = 0 \text{ for all tasks } T_i.$$

A policy σ is called *optimal* if for any feasible power profile p, the power allocations under σ also complete the tasks. This notion of optimality is based only on power profile feasibility and is *unrelated* to the cost metric (2).

We seek causal CM scheduling policies that are optimal - policies that satisfy task requirements for *all* feasible power profiles using only current information. In addition, we want these policies to meet task requirements at low cost as captured by the metric (2).

III. Main Results

We first show that *causal* optimal scheduling policies which respect servicing power limits *do not exist*. We then present three heuristic causal scheduling policies: Earliest Deadline First [EDF], Least Laxity First [LLF], and Receding Horizon Control [RHC] to allocate available generation to tasks. We show that EDF is optimal in the absence of rate constraints.

A. Causal Optimal Policies Do Not Exist

We prove that there is no causal optimal scheduling policy for the resource scheduling problem using an adversarial argument.

Theorem 1: There does not exist an causal optimal scheduling policy.

Proof: By counterexample. Consider the following tasks:

Task 1: $E_1 = 2$, $a_1 = 0$, $d_1 = 2$, $m_1 = 2$.
Task 2: $E_2 = 2$, $a_2 = 0$, $d_2 = 4$, $m_2 = 1$.

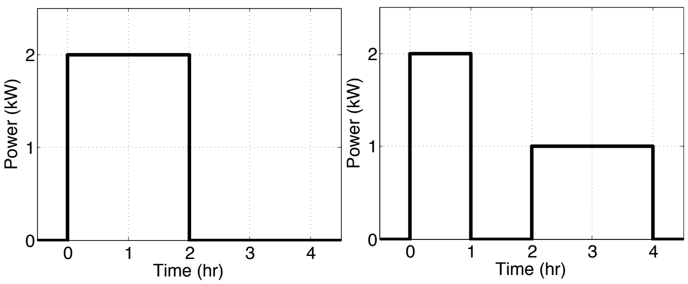
Consider the two available power profiles p^A and p^B shown in Figure 2. Both p^A and p^B are feasible. We exhibit the associated power allocations that meet task constraints and complete the tasks in Figure 3. We first argue that these power allocations are *unique*. For profile p^A, completion of task 2 requires 2 units of energy at a maximum rate of 1. Since p^A has support $[0, 2]$, we must service task 2 at its maximum rate. The remaining power $p^A - p_2^A$ must be allocated entirely to task 1. For profile p^B, completion of task 1 requires 2 units of energy at a maximum rate of 2. This task must be serviced on the interval $[0, 2]$. Since $p^B = 0$ on $[1, 2]$, task 1 must be serviced on $[0, 1]$ at its maximum rate. The remaining power $p^B - p_1^B$ has exactly 2 units of energy and must be allocated entirely to task 2.

Next notice that the two available power profiles p^A, p^B are *identical* on $t \in [0, 1]$. Therefore, any *causal* policy σ must offer identical allocations under either power profile on $t \in [0, 1]$. If σ is optimal, it must complete both tasks under p^A *and* p^B. But completion of both tasks under these feasible available power profiles requires *different* allocations for $t \in [0, 1]$. ■

Since we cannot, in general, construct *causal* optimal scheduling policies, we must be content with sub-optimal heuristic scheduling algorithms.

B. Earliest Deadline First [EDF]

Earliest deadline first [EDF] is a well-studied scheduling algorithm in the context of processor time allocation [PTA]



(a) Profile p^A (b) Profile p^B

Fig. 2: Two profiles that show that causal optimal scheduling policies do not exist.

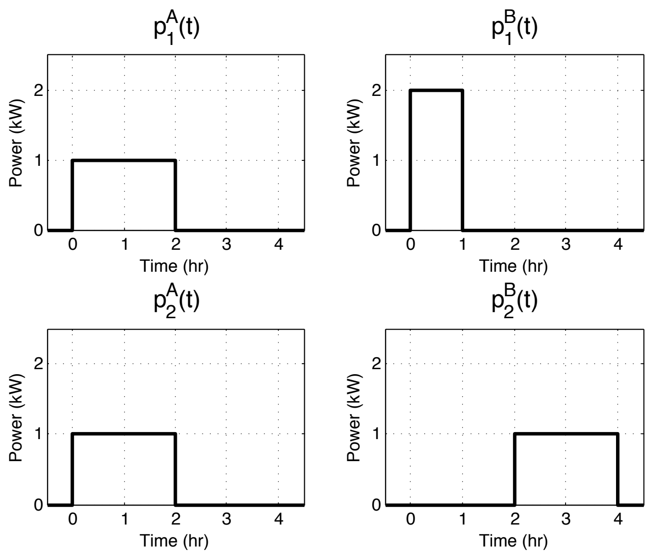


Fig. 3: Power supplied to tasks 1 and 2 according to feasible scheduling policies for Profiles p^A and p^B

[1], [2]. The objective in PTA is to schedule a collection of computation tasks on a single processor. This is analogous to the scheduling problem we consider in this paper with processor capacity [i.e. speed] playing the role of available power. In both cases, the objective is to design schedules that complete a collection of tasks.

The EDF scheduling policy first assigns available power $p(t)$ to the active task T_j with the earliest deadline:

$$j(t) = \operatorname*{argmin}_{i \in \mathbb{A}_t} d_i. \tag{3}$$

Available generation $p(t)$ in excess of the rate limit for task T_j is allocated to the active task with the next most imminent deadline. Ties are broken arbitrarily. This process continues until all available power $p(t)$ is expended.

For the single processor case, it is well known that EDF is an optimal scheduling policy [1]. Specifically, if there exists some scheduling policy that completes all tasks, the scheduling policy under EDF also completes all tasks. Unfortunately, this result does not apply to the resource scheduling problem for three reasons. First, while processor capacity is fixed and constant over the entire interval of operation, its analog in resource scheduling, available generation, is variable. Second, rate constraints effectively limit the amount of power supplied to a particular task at a given point in time, a consideration absent in PTA. Third, at any time, a processor works solely on, and devotes all its processing capacity to a single task. This is not the case in resource scheduling. Multiple tasks may be scheduled concurrently at any time, and each task may receive power at a rate bounded by its servicing power limit m_i.

We first show that EDF is optimal when there are *no rate constraints*. The proof technique used here closely follows that presented in [1] for PTA, except that we focus on scheduling available energy as opposed to available processor time. This is required to account for concurrent task scheduling, a major difference between the PTA and resource scheduling problems.

Theorem 2 (EDF Optimality): If $p(t)$ is a feasible aggregate power profile for a collection of tasks $\{T_i\}_{i=1}^M$, then the scheduling policy defined by EDF satisfies all load energy requirements in the *absence of rate constraints*.

Proof: Let σ be a scheduling policy for $p(t)$ that completes a collection of tasks $\{T_i\}_{i=1}^M$ and σ_{EDF} be the EDF scheduling policy for the same power profile. The proof technique involves perturbing energy allocations made by σ to show that σ_{EDF} also completes all tasks.

If σ and σ_{EDF} are identical, the claim follows trivially. Assume these policies are distinct. Let t_0 be the first time at which σ and σ_{EDF} differ. Over a small time interval $[t_0, t_0 + \delta t_0]$, assume σ assigns energy δE to task T_1 while σ_{EDF} assigns energy δE to task T_2.

Since σ completes all tasks, it must satisfy the energy requirement for task T_2. Thus, there exists some time interval $[t_1, t_1 + \delta t_1]$ before the deadline for T_2 over which σ assigns δE to task T_2. Create a new scheduling policy $\hat{\sigma}$ that is identical to σ except T_2 is scheduled over $[t_0, t_0 + \delta t_0]$ and T_1 is scheduled over $[t_1, t_1 + \delta t_1]$. Clearly, $\hat{\sigma}$ completes all tasks as $t_1 + \delta t_1 \le d_2 \le d_1$. The last inequality follows from the task choice of σ_{EDF} at time t_0. This procedure can be applied iteratively, with $\hat{\sigma}$ replacing σ on each iteration, until the resulting policy is σ_{EDF}, proving the claim. ■

This result does not apply to general resource power allocation for two reasons. First, the CM scheduling problem [Section II] involves task rate constraints. Second, without advance knowledge of the renewable generation realization and all task parameters, it is impossible to *a priori* certify the feasibility of an available power profile $p(t)$. Nonetheless, our simulation studies [Section IV] suggest that an EDF-based renewable generation scheduling policy coupled with a suitable policy for grid generation offers a reasonable heuristic for real-time resource scheduling.

C. Least Laxity First [LLF]

EDF schedules tasks solely according to their deadlines and does not take into account task energy states. An alternative to EDF is least laxity first [LLF], another well-known algorithm in the context of PTA [3]. LLF takes into account task energy states and has been studied as a scheduling policy for EV charging [22].

Let $e_i(t)$ be the energy state of task T_i. Define the *deferrable deadline* for the task T_i,

$$\delta_i(t) = d_i - \frac{e_i(t)}{m_i}.$$

Note that $\delta_i(t)$ is the latest time at which task T_i can be initiated and successfully completed.

The LLF scheduling policy first assigns available power $p(t)$ to the active task T_j with earliest deferrable deadline:

$$j(t) = \operatorname*{argmin}_{i \in \mathbb{A}_t} \delta_i(t). \tag{4}$$

Available generation $p(t)$ in excess of the rate limit for task T_j is allocated to the active task with the next most recent deferrable deadline. Ties are broken arbitrarily. This process continues until all available power $p(t)$ is expended.

We use the deferrable deadline to quantify the degree of flexibility in scheduling task T_i at time t. Define the *flexibility factor* for the task T_i as:

$$\phi_i(t) = (d_i - t) - \frac{e_i(t)}{m_i} = \delta_i(t) - t. \tag{5}$$

This is the difference between the amount of time remaining to complete the task and the time required to complete it at its maximum rate. Larger flexibility factors imply greater load deferability. In particular, if task T_i is not flexible $[\phi_i(t) = 0]$, it *must* be serviced immediately at rate m_i to be completed by its deadline. In the LLF scheduling policy, the CM schedules least flexible tasks first.

We remark that both EDF and LLF are extremely easy to implement. While resources can game such policies by publishing false deadlines, pricing energy according to flexibility factors disincentivizes such strategic behavior.

D. Grid Generation Scheduling Policy

Flexibility factors can be used to formulate grid generation scheduling policies. Clearly, a task cannot be completed by its deadline if the minimum time required to satisfy the energy requirement exceeds the time remaining in the service interval. Specifically, a task T_i is infeasible at time t if $\phi_i(t) < 0$.

An optimal grid power scheduling policy would need knowledge of *future* renewable generation scheduling decisions. As we restrict ourselves to causal scheduling policies, a reasonable heuristic is as follows. Schedule grid power for active tasks with flexibility factors within a threshold $\epsilon > 0$. Procure adequate grid power to ensure the total power allocated to each such task T_i from both renewable and grid generation is m_i.

We submit that EDF or LLF policies for scheduling renewable generation and this flexibility factor-based policy for scheduling grid generation offers a reasonable causal heuristic for the CM resource allocation problem that is computationally attractive.

E. Receding Horizon Control [RHC]

We now present a RHC approach to solving the resource scheduling problem. This is a discrete-time control strategy inspired by model predictive control [MPC] [7]. We remark that several authors have considered MPC/RHC strategies for EV charging in particular, and various scheduling problems that arise in power systems operation in general [9], [19], [26], [27]. Our contribution is not in suggesting the use of RHC methods, but rather, in the particular choice of objective function [inspired by PTA methods] that we make.

Let the horizon length N be the number of Δt-length time-steps between t and the final deadline in the set of active tasks $\mathbb{A}_t$. Note that N varies with time t. Let M be the number of active tasks. Each optimization problem attempts to allocate forecasted generation $\hat{w}_k$ to complete all M tasks over the time horizon $[t, t + N\Delta t]$.

1) Decision Variables: Let the amount of energy delivered to task i at time $t + k\Delta t$ by renewable and grid generation be W_{ik} and G_{ik} respectively. Define the matrices $W, G \in \mathbb{R}^{M \times N}$ whose i, k entries are W_{ik} and G_{ik}.

2) Objective function: We propose the objective function:

$$\begin{aligned} J(W,G) = \quad & \alpha_1 \|\mathbf{1}^T G\|_1 + \alpha_2 \|\mathbf{1}^T G\|_\infty \\ & + \sum_{i \in \mathbb{A}_t} \sum_{k=1}^{N} (N - \phi_i(k))^2 \end{aligned} \tag{6}$$

The first two terms in (7) capture grid energy, and grid capacity costs respectively. The third term in (7) helps maintain adequate task flexibility for all active tasks across the time horizon. This incentivizes earlier allocations of as much renewable generation as possible. Effectively, this term penalizes unused renewable generation [down regulation]. The parameters α_1 and α_2 negotiate the relative importance of the objective function components.

3) Problem Statement: The optimization problem solved at each time t is:

$$\min_{W,G} J(W,G)$$

$$\text{s.t.: } W^T \mathbf{1} \le \hat{w} = [\hat{w}_1 \hat{w}_2 \ldots \hat{w}_N]^T \tag{7}$$

$$(W + G)\,\mathbf{1} = E = [E_1 E_2 \ldots E_M]^T \tag{8}$$

$$\forall k, W_{ik} + G_{ik} \begin{cases} \in [0, m_i \Delta t]\,, & \forall i : t + k\Delta t \le d_i \\ = 0, & \forall i : t + k\Delta t > d_i \end{cases} \tag{9}$$

$$\phi_i(k) = d_i - (t + k\Delta t) - \frac{e_i(k)}{m_i} \tag{10}$$

$$e_i(k) = E_i - \sum_{j=1}^{k} W_{ij} + G_{ij} \tag{11}$$

4) Constraints:

(a) *Generation*: The sum of all power allocated to tasks cannot exceed the forecasted renewable generation [$\hat{w}_k$] at any time $t + k\Delta t$ (7).
(b) *Energy Requirement*: Each task's energy requirement [E_i] is met through allocation of renewable and grid generation (8).
(c) *Rate Limits*: The power dispatched to an active task at a time within its service interval is non-negative and bounded by its rate limit [m_i] (9).
(d) *Flexibility factors*: The constraints (10) and (11) are used to compute the factors $\phi_i(k)$.

IV. Examples

In this section, we illustrate the reserve cost reduction afforded by coordinated resource aggregation using simulated test case scenarios to satisfy the energy needs of 100 EVs. Power from representative wind power trajectories is allocated according to the various heuristics offered in Section III. We then compare the simulation results on the basis of grid energy and capacity requirements.

A. Simulation Parameters

In each simulation, renewable and grid generation scheduling decisions are made every 5 minutes over a 12 hour operating window. Arrival times, deadlines, and energy requirements are randomly chosen for each scheduling task based on typical EV charging specifications [23]. These parameters, while random, are chosen to ensure each task is feasible, i.e. $E_i \leq m_i(d_i - a_i)$. We also enforce identical maximum rate of charge [m_i] for all tasks. This is a reasonable assumption as EV battery characteristics and distribution network constraints are similar across tasks.

We use wind power time series data, sampled every 5 minutes, from the Bonneville Power Administration [BPA] to generate sample renewable generation profiles. These profiles are rescaled so that the total energy generated from renewables is equal to the total load energy requirement. Since the objective of this study is to analyze the effectiveness of different scheduling policies at mitigating renewable generation variability, we assume this wind power profile represents all available renewable generation. Any additional power necessary to meet load requirements can only be obtained via reserves [grid generation].

We synthetically create renewable generation forecasts for the RHC controller by adding Gaussian noise to the wind power profiles. Forecasts $\hat{w}_k$ are generated according to:

$$\hat{w}_k = w_k + \sum_{n=1}^{k-1} \epsilon_n, k = 2, 3, \ldots, N \tag{12}$$
$$\epsilon_n \sim \mathcal{N}(0, \sigma_n^2), \hat{g}_1 = g_1.$$

where σ_n^2 is an increasing function of time [indexed by n] as forecast uncertainty increases with prediction horizon length. As a result, the deviation between the forecasted and actual wind power profiles increases with time. Figure 4 shows a 12-hour wind profile and a sample forecast for this profile.

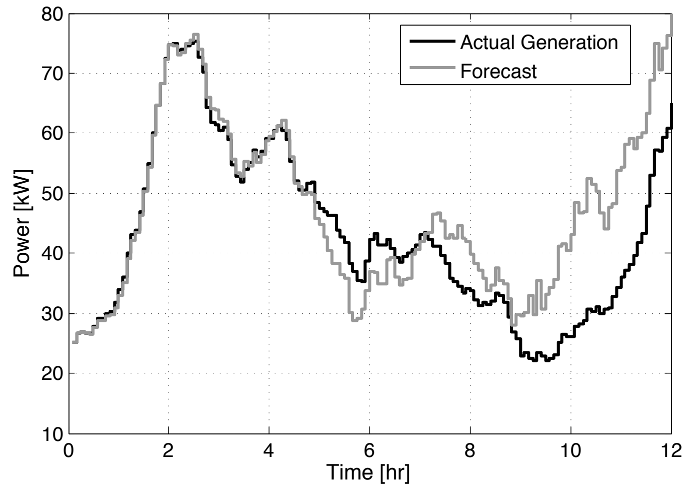


Fig. 4: Sample forecast generated for a 12 hr wind profile

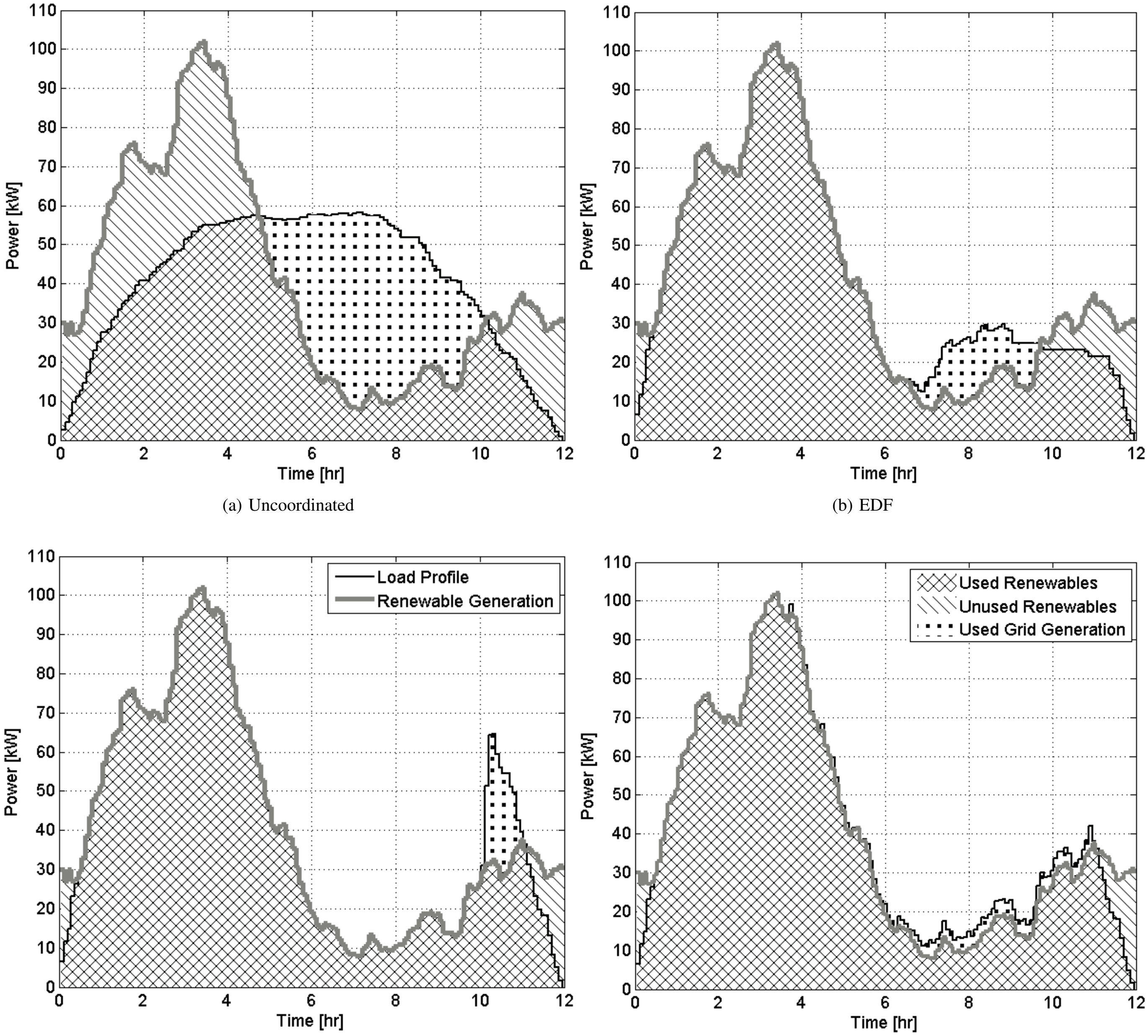
B. Comparison

Figure 5 shows the amounts of renewable and grid generation allocated to meet load requirements over time under the various scheduling policies. Each subfigure has three shaded areas which correspond to:

(a) Renewable generation dispatched to tasks.
(b) Unused renewable generation.
(c) Grid generation dispatched to tasks.

The generation and load profiles are shown in Figure 5. Figure 5a corresponds to the case with no resource scheduling. In this case, each task receives a constant power profile over its service interval that satisfies its energy requirement. The value of coordinated scheduling is immediately apparent from comparing Figures 5b, 5c, and 5d to Figure 5a. Resource scheduling results in aggregate load profiles that better approximate the renewable generation profile. This clearly suggests a CM can mitigate some of the variability associated with renewable generation through judicious allocation of power to deferrable loads.

When employing EDF- or LLF-based scheduling [Figures 5b and 5c], grid generation is procured only toward the end of task service intervals. Under these scheduling policies, the CM allocates reserve generation only when there is no longer any flexibility in task scheduling. This phenomenon tends to occur close to task deadlines. Further, with LLF-based scheduling, the flexibility factors for all active tasks are the *same* when the CM first schedules grid generation. This explains the large spikes in grid power acquisition in Figure 5c. In contrast, grid power acquisition under RHC is far more balanced [Figure 5d] as the CM incorporates generation forecasts in procurement decisions.



(a) Uncoordinated (b) EDF (c) LLF (d) RHC

Fig. 5: Generation and reserve scheduling for deferrable loads with no coordinated and under EDF, LLF, and RHC

In order to quantify the effectiveness of a scheduling policy, we focus on three quantities: renewable generation used, grid energy required, and grid capacity required. The grid capacity requirement is calculated by finding the maximum instantaneous usage of grid power. We compute these quantities for each scheduling policy over 100 different simulated test cases. Table I shows average percentage decreases in these quantities over all test cases. It is clear that *any* form of scheduling increases the amount of renewable generation used to serve loads and thus, reduces the need to dispatch grid power. Moreover, coordinated scheduling under any policy results in at least a 50% decrease in dispatched grid energy as compared to that in the case without coordination. The particular choice of scheduling policy has little impact on this performance metric. Conversely, the grid capacity requirement is highly dependent on the choice of scheduling policy. RHC-based scheduling results in a 60% reduction in required reserve capacity while LLF-based scheduling requires a 70% increase in reserve capacity than is needed without coordinated scheduling.

Coordinated scheduling of deferrable loads decreases grid energy costs for all three policies. Grid capacity costs de-

	EDF	LLF	RHC
Renewable energy used	-24.87	-26.94	-27.16
Grid energy required	55.78	62.79	63.02
Grid capacity required	12.81	-70.90	66.99

TABLE I: Percentage decreases in metrics for assessing scheduling policy performance

crease for EDF- and RHC-based scheduling, with RHC-based scheduling achieving the lowest cost. This is not surprising as RHC-based scheduling uses generation forecasts to find optimal power allocations while EDF- and LLF-based scheduling do not. RHC, however is more computationally intensive requiring convex optimization.

V. Conclusion

In this paper, we have investigated real-time power scheduling to deferrable loads. We have shown that optimal causal scheduling policies do not exist. We then explored three scheduling algorithms for renewable generation: EDF, LLF, and RHC, and a flexibility-based heuristic for grid generation [reserves]. We quantified the relative effectiveness of these algorithms through simulated test cases. We showed, through these simulations, that coordinated scheduling via any of these 3 policies decreases the required reserve energy to meet load requirements while only EDF and RHC reduce the reserve capacity requirement.

There are two important topics that have not been explored in this paper, and will be the subject of our future research:

(a) *Architecture and implementation*. Implementation of various cluster manager algorithms [ex: renewable and grid generation scheduling] requires an underlying communication/computation architecture. There are two options: centralized or [partially] distributed architectures. The architectural choice will depend on the performance requirements [ex: latency, flops, reliability] needed. These essential aspects require exploration.

(b) *Reserve reduction*. Our case for coordinated resource aggregation rests on the improved forecastability it offers in *ex-ante* markets. This, in turn, will require less reserve capacity requirements for load following. The economic basis for this claim requires justification through synthetic examples and much more comprehensive simulations using real data.

References

[1] S. Baruah, and J. Goossens, "Scheduling real-time tasks: Algorithms and Complexity," in *Handbook of Scheduling: Algorithms, Models and Performance Analysis,* J.Y-T. Leung, Ed. Boca Raton, FL: CRC Press, 2004, chpt. 28.

[2] C.L. Liu, and J.W. Layland, "Scheduling algorithms for multiprogamming in a hard-real-time environment," *Journal of the ACM*, 20(1): 46-61, January 1973.

[3] M.L. Dertouzos, and A.K. Mok, "Multiprocessor online scheduling of hard-real-time tasks," *IEEE Trans. Softw. Eng.*, 15(12): 1497-1506, December 1989.

[4] M. Alizadeh, A. Scaglione, R.J. Thomas, and D. Callaway, "Information infrastructure for cellular load management in green power delivery systems," *IEEE SmartGridComm Conf.*, 2011.

[5] E. Bitar, P. Khargonekar, K. Poolla, R. Rajagopal, P. Varaiya, and F. Wu, "Selling random wind," *45th Hawaii International Conference on System Science (HICSS)*, 2012.

[6] E. Bitar, R. Rajagopal, P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," submitted to *IEEE Trans. Power Syst.*, 2011.

[7] F. Borrelli, A. Bemporad, M. Morari, "Constrained optimal control and predictive control for linear and hybrid systems," in press. [Online]. Available: http://www.mpc.berkeley.edu.

[8] California Independent System Operator (CAISO), "Integration of renewable resources: Operational requirements and Generation fleet capability at 20% RPS," Technical Report, August 2010.

[9] M. Galus, R. la Fauci, and G. Andersson, "Investigating PHEV wind balancing capabilities using heuristics and model predictive control", *IEEE Power & Energy Society General Meeting*, 2010.

[10] M. Roozbehani, M. Dahleh, and S.K. Mitter, "On the stability of wholesale electricity markets under real-time pricing", *49th Conf. on Decision and Control*, 2010.

[11] D.S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, 50(5): 1389-1400, May 2009.

[12] A.J. Conejo, J.M. Morales, and J.A. Martinez, "Tools for the analysis and design of distributed resources - Part III: market studies," *IEEE Trans. Power Del.*, 26(3): 1663-1670, July 2011.

[13] D.J. Hammerstrom, J. Brous, D.P. Chassin, et al., "Pacific Northwest GridWise testbed demonstration projects," Tech. Rep. PNNL-17167, Pacific Northwest National Laboratory, October 2007.

[14] G. Heffner, C. Goldman, B. Kirby, and M. Kintner-Meyer, "Loads providing ancillary services: Review of international experience," Technical Report LBNL-62701, Lawrence Berkeley National Laboratory, May 2007.

[15] U. Helman, "Resource and transmission planning to achieve a 33% RPS in California - ISO Modeling tools and planning framework," *FERC Technical Conference on Planning Models and Software*, 2010.

[16] Y. Y. Hsu, C. C. Su, "Dispatch of direct load control using dynamic programming," *IEEE Trans. Power Syst.*, 6(3): 1056-1061, August 1991.

[17] L. Jiang, and S. Low, "Multi-period optimal procurement and demand responses in the presence of uncertain supply," *50th Conf. on Decision and Control*, December 2011.

[18] T.F. Lee, M.Y. Cho, Y.C. Hsiao, P.J. Chao, and F.M. Fang, "Optimization and implementation of a load control scheduler using relaxed dynamic programming for large air conditioner loads," *IEEE Trans. Power Syst.*, 23(2): 691-702, May 2008.

[19] M. Ilic, L. Xie, and J.Y. Joo, "Efficient coordination of wind power and price-responsive demand", *IEEE Trans. Power Syst.*, 26(4): 1885-1893, November 2011.

[20] North American Electric Reliability Corporation (NERC),"Accommodating high levels of variable generation," White Paper, April 2009.

[21] A. Papavasiliou, and S. Oren, "Supplying renewable energy to deferrable loads: Algorithms and economic analysis," *IEEE Power & Energy Society General Meeting*, 2010.

[22] S. Chen, T. He, and L. Tong, "Optimal deadline scheduling with commitment", *Allerton Conference on Communication, Control, and Computing*, 2011.

[23] S.Shao, M. Pipattanasomporn, and S. Rahman, "Challenges of PHEV penetration to the residential distribution network," *IEEE Power & Energy Society General Meeting*, 2009.

[24] J.A. Short, D.G. Infield, and L.L. Freris, "Stabilization of grid frequency through dynamic demand control," *IEEE Trans. Power Syst.*, 22(3): 1284-1293, August 2007.

[25] K. Spees and L.B. Lave, "Demand response and electricity market efficiency," *The Electricity Journal*, 20(3): 69-85, April 2007.

[26] Andreas Ulbig, Michèle Arnold, Spyros Chatzivasileiadis, and Göran Andersson, "Framework for multiple time-scale cascaded MPC application in power systems", *18th IFAC World Congr.*, 2011.

[27] G. Hug-Glanzmann, "Coordination of intermittent generation with storage, demand control and conventional energy sources", *IREP 2010 - Bulk Power System Dynamics and Control - VIII*, 2010.