Forty-Ninth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 28 - 30, 2011

Distributed Storage for Intermittent Energy Sources: Control Design and Performance Limits

Yashodhan Kanoria, Andrea Montanari, Stanford University
David Tse and Baosen Zhang, UC Berkeley

***Abstract*— One of the most important challenges in the integration of renewable energy sources into the power grid lies in their 'intermittent' nature. The power output of sources like wind and solar varies with time and location due to factors that cannot be controlled by the provider. Two strategies have been proposed to hedge against this variability: 1) use energy storage systems to effectively average the produced power over time; 2) exploit distributed generation to effectively average production over location. We introduce a network model to study the optimal use of storage and transmission resources in the presence of random energy sources. We propose a Linear-Quadratic based methodology to design control strategies, and show that these strategies are asymptotically optimal for some simple network topologies. For these topologies, the dependence of optimal performance on storage and transmission capacity is explicitly quantified.**

I. Introduction

It is widely advocated that future power grids should facilitate the integration of a significant amount of renewable energy sources. Prominent examples of renewable sources are wind and solar. These differ substantially from traditional sources in terms of two important qualitative features:

Intrinsically distributed. The power generated by these sources is typically proportional to the surface occupied by the corresponding generators. For instance, the solar power reaching ground is of the order of 2 kWh per day per square meter. The wind power at ground level is of the order of 0.1 kWh per day per square meter [1]. These constraints on renewable power generation have important engineering implications. If a significant part of energy generation is to be covered by renewables, generation is argued to be distributed over large geographical areas.

Intermittent. The output of renewable sources varies with time and locations because of exogenous factors. For instance, in the case of wind and solar energy, the power output is ultimately determined by meteorological conditions. One can roughly distinguish two sources of variability: *predictable variability*, e.g. related to the day-night cycle, or to seasonal differences; *unpredictable variability*, which is most conveniently modeled as a random process.

Several ideas have been put forth to meet the challenges posed by intermittent production. The first one is to leverage geographically distributed production. The output of distinct generators is likely to be independent or weakly dependent over large distances and therefore the total production of a large number of well separated generators should stay approximately constant, by a law-of-large-number effect.

The second approach is to use energy storage to take advantage of over-production at favorable times, and cope with shortages at unfavorable times. Finally, a third idea is 'demand response', which aims at scheduling in optimal ways some time-insensitive energy demands. In several cases, this can be abstracted as some special form of energy storage (for instance, when energy is demanded for interior heating, deferring a demand is equivalent to exploiting the energy stored as hot air inside the building).

These approaches hedge against the energy source variability by averaging over location, or by averaging over time. Each of them requires specific infrastructures: a power grid with sufficient transmission capacity in the first case, and sufficient energy storage infrastructure in the second one. Further, these two directions are in fact intimately related. With current technologies, it is unlikely that centralized energy storage can provide effective time averaging of –say– wind power production, in a renewables-dominated scenario. In a more realistic scheme, storage is distributed at the consumer level, for instance leveraging electric car batteries (a scenario known as vehicle-to-grid or V2G). Distributed storage implies, in turn, substantial changes of the demand on the transmission system.

The use of storage devices to average out intermittent renewables production is well established. A substantial research effort has been devoted to its design, analysis and optimization, see for instance [2], [3], [4], [5], [6], [7]. In this line of work, a large renewable power generator is typically coupled with a storage system in order to average out its power production. Proper sizing, response time, and efficiency of the storage system are the key concerns.

If, however, we assume that storage will be mainly distributed, the key design questions change. It is easy to understand that both storage and transmission capacity will have a significant effect on the ability of the network to average out the energy source variability. For example, shortfalls at a node can be compensated by either withdrawals from local storage or extracting power from the rest of the network, or a combination of both. The main goal of this paper is to understand the optimal way of utilizing simultaneously these two resources and to quantify the impact of these two resources on performance. Our contributions are:

- a simple model capturing key features of the problem;
- a Linear-Quadratic(LQ) based methodology for the systematic design of control strategies;
- a proof of optimality of the LQ control strategies in simple network topologies such as the 1-D and 2-D grids and in certain asymptotic regimes.
- a quantification of how the performance depends on key

parameters such as storage and transmission capacities. The reader interested in getting an overview of the conclusions without the technical details can read Sections II and IV only. Some details are omitted due to space limitations and a full version is posted on ArXiv.

II. Model and Problem Formulation

The power grid is modeled as a weighted graph G with vertices (buses or nodes) V, edges (lines) E. Time is slotted and will be indexed by $t \in \{0, 1, 2, \dots\}$. In slot t, each node $i \in V$ receives a supply of a quantity of energy $E_{\mathrm{p},i}(t)$ from a renewable source, and receives a demand of a quantity $E_{\mathrm{d},i}(t)$ for consumption. For our purposes, these quantities only enter the analysis through the net supply $\mathcal{Z}_i(t) = E_{\mathrm{p},i}(t) - E_{\mathrm{d},i}(t)$. Let $\mathcal{Z}(t$ be the vector of $\mathcal{Z}_i(t)$'s. We will assume that $\{\mathcal{Z}(t)\}$ is a stationary process.

In order to average the variability in the energy supply, the system makes use of storage and transmission. Storage is fully distributed: each node $i \in V$ has a device that can store energy, with capacity S_i. We assume that stored energy can be fully recovered when needed (i.e., no losses). At each time slot t, one can transfer an amount of energy $\mathcal{Y}_i(t)$ to storage at node i. If we denote by $\mathcal{B}_i(t)$ the amount of stored energy at node i just before the beginning of time slot t, then:

$$B_i(0) = 0, \qquad \mathcal{B}_i(t+1) = [\mathcal{B}_i(t) + \mathcal{Y}_i(t)]_0^{S_i} \tag{1}$$

where $[x]_a^b := \max(\min(x, b), a)$ for $a \leq b$.

We will also assume the availability at each node of a fast generation source (such as a spinning reserve or backup generator) which allows covering up of shortfalls. Let $\mathcal{W}_i(t)$ be the energy obtained from such a source at node i at time slot t. We will use the convention that $\mathcal{W}_i(t)$ is negative means that energy is consumed from the fast generation source, and positive means energy is dumped. The cost of using fast generation energy sources is reflected in the steady-state performance measure:

$$\varepsilon_{\mathcal{W}} \equiv \lim_{t\to\infty} \frac{1}{|V|} \sum_{i \in V} \mathbb{E}\{\big(\mathcal{W}_i(t)\big)_-\} \tag{2}$$

The net amount of energy injection at node i at time slot t is:

$$\mathcal{Z}_i(t) - \mathcal{Y}_i(t) - \mathcal{W}_i(t).$$

These injections have to be distributed across the transmission network, and the ability of the network to distribute the injections and hence to average the random energy sources over space is limited by the transmission capacity of the network. To understand this constraint, we need to relate the injections to the power flows on the transmission lines. To this end, we adopt a 'DC power flow' approximation model [8]. [1]

Each edge in the network corresponds to a transmission line which is purely inductive, i.e. with conductance $-jb_e$, where $b_e \in \mathbb{R}_+$. Hence, the network is lossless. Node $i \in V$ is at voltage $V_i(t)$, with all the voltages assumed to have the same magnitude, taken to be 1 (by an appropriate choice of units). Let $V_i(t) = e^{j\phi_{i,t}}$ denote the (complex) voltage at node i in time slot t. If $I_{i,k}(t) = -jb_{ik}(V_i(t) - U_k(t))$ is the electric current from i to k, the corresponding power flow is then $\mathcal{F}_{i,k}(t) = \mathrm{Re}[V_i(t) I_{i,k}(t)^*] = \mathrm{Re}[jb_{ik}(1 - e^{j(\phi_i(t) - \phi_k(t))})] = b_{ik} \sin(\phi_i(t) - \phi_k(t))$, where $\mathrm{Re}[\cdot]$ denotes the real part of a complex number.

The DC flow approximation replaces $\sin(\phi_i(t) - \phi_k(t))$ by $\phi_i(t) - \phi_k(t)$ in the above expression. This is usually a good approximation since the phase angles at neighboring nodes are typically maintained close to each other to ensure that the generators at the two ends remain in step. This leads to the following relation between angles and power flow $\mathcal{F}_{i,k}(t) = b_{ik}(\phi_i(t) - \phi_k(t))$. In matrix notation, we have

$$\mathcal{F}(t) = \nabla \phi(t), \tag{3}$$

where $\mathcal{F}(t)$ is the vector of all power flows, $\phi(t) = (\phi_1(t), \dots, \phi_n(t))$ and ∇ is a $|E| \times |V|$ matrix. $\nabla_{e,i} = b_e$ if $e = (i,k)$ for some k, $\nabla_{e,i} = -b_e$ if $e = (k,i)$ for some k, and $\nabla_{e,i} = 0$ otherwise.

Energy conservation at node i also yields

$$\mathcal{Z}_i(t) - \mathcal{Y}_i(t) - \mathcal{W}_i(t) = \sum_k \mathcal{F}_{(i,k)}(t) = \big(\nabla^T \mathbf{b}^{-1} F(t)\big)_i,$$

where $\mathbf{b} = \mathrm{diag}(b_e)$ is an $|E| \times |E|$ diagonal matrix. Expressing $\mathcal{F}(t)$ in terms of $\phi(t)$, we get

$$\mathcal{Z}(t) - \mathcal{Y}(t) - \mathcal{W}(t) = -\Delta \phi(t), \tag{4}$$

where $\Delta = -\nabla^T \mathbf{b}^{-1} \nabla$ is a $|V| \times |V|$ symmetric matrix where $\Delta_{i,k} = -\sum_{l:(i,l)\in E} b_{il}$ if $i = k$, $Delta_{i,k} = b_{ik}$ if $(i,k) \in E$ and 0 otherwise. In graph theory, Δ is called the graph Laplacian matrix. In power engineering, it is simply the imaginary part of the bus admittance matrix of the network. Note that if $b_e \geq 0$ for all edges e, then $-\Delta \succcurlyeq 0$ is positive semidefinite. If the network is connected (which we assume throughout), it has only one eigenvector with eigenvalue 0, namely the vector $\varphi_v = 1$ everywhere. This fits the physical fact that if all phases are rotated by the same amount, the powers in the network are not changed.

With an abuse of notation, we denote by Δ^{-1} the matrix that has the same kernel as Δ, and is equal to the inverse of Δ on the orthogonal subspace. Explicitly $\Delta = -\mathbf{V}\alpha^2\mathbf{V}^T$ be the eigenvalue decomposition of Δ, where α is a diagonal matrix with non-negative entries. Define $\alpha^\dagger$ to be the diagonal matrix with $\alpha^\dagger_{ii} = 0$ if $\alpha_{ii} = 0$ and $\alpha^\dagger_{ii} = \alpha^{-1}_{ii}$ otherwise. Then $\Delta^{-1} = -\mathbf{V}(\alpha^\dagger)^2\mathbf{V}^T$.

Since the total power injection in the network adds up to zero (which must be true by energy conservation), we can invert (4) and obtain

$$\phi(t) = -\Delta^{-1}(\mathcal{Z}(t) - \mathcal{Y}(t) - \mathcal{W}(t)). \tag{5}$$

Plugging this into (3), we have

$$\mathcal{F}(t) = -\nabla\Delta^{-1}\big(\mathcal{Z}(t) - \mathcal{Y}(t) - \mathcal{W}(t)\big). \tag{6}$$

There is a capacity limit C_e on the power flow along each edge e; this capacity limit depends on the voltage

[1] Despite the name, 'DC flow' is an approximation to the AC flow

magnitudes and the maximum allowable phase differences between adjacent nodes, as well as possible thermal line limits. We will measure violations of this limit by defining

$$\varepsilon_{\mathcal{F}} \equiv \lim_{t\to\infty} \frac{1}{|E|} \sum_{e\in E} \mathbb{E}\{(\mathcal{F}_e(t) - C_e)_+ + (-C_e - \mathcal{F}_e(t))_+\}\,. \tag{7}$$

We are now ready to state the design problem:

For the dynamic system defined by equations (1) and (6), design a control strategy which, given the past and present random renewable supplies and the storage states,

$$\big\{(\mathcal{Z}(t), \mathcal{B}(t));\ (\mathcal{Z}(t-1), \mathcal{B}(t-1)), \dots, \big\}$$

choose the vector of energies $\mathcal{Y}(t)$ to put in storage and the vector of fast generations $\mathcal{W}(t)$ such that the sum $\varepsilon_{\rm tot} \equiv \varepsilon_{\mathcal{F}} + \varepsilon_{\mathcal{W}}$, cf. Eq. (2) and (7), is minimized.

III. Linear-Quadratic Design

In this section, we propose a design methodology that is based on Linear-Quadratic (LQ) control theory.

A. The Surrogate LQ Problem

The difficulty of the control problem defined above stems from both the nonlinearity of the dynamics due to the hard storage limits and the piecewise linearity of the cost functions giving rise to the performance parameters. Instead of attacking the problem directly, we consider a surrogate LQ problem where the hard storage limits are removed and the cost functions are quadratic:

$$B_i(0) = \frac{-S_i}{2}, \qquad B_i(t+1) = B_i(t) + Y_i(t) \tag{8}$$

$$F(t) = -\nabla\Delta^{-1}\big(Z(t) - Y(t) - W(t)\big) \tag{9}$$

with performance parameters:

$$\varepsilon^{\rm surrogate}_{W_i} = \lim_{t\to\infty} \mathbb{E}\{\big(W_i(t)\big)^2\}, \quad i \in V \tag{10}$$

$$\varepsilon^{\rm surrogate}_{F_e} = \lim_{t\to\infty} \mathbb{E}\{(F_e(t))^2\}, \quad e \in E \tag{11}$$

$$\varepsilon^{\rm surrogate}_{B_i} = \lim_{t\to\infty} \mathbb{E}\{(B_i(t))^2\} \tag{12}$$

The process $B_i(t)$ can be interpreted as the deviation of a virtual storage level process from the midpoint $S_i/2$, where the virtual storage level process is no longer hard-limited but evolves linearly. Instead, we penalize the deviation through a quadratic cost function in the additional performance parameters $\varepsilon^{\rm surrogate}_{B_i}$.

The virtual processes $B(t)$, $F(t)$, $W(t)$, $Y(t)$ and $Z(t)$ are connected to the actual processes $\mathcal{B}(t)$, $\mathcal{F}(t)$, $\mathcal{W}(t)$, $\mathcal{Y}(t)$ and $\mathcal{Z}(t)$ via the mapping:

$$\mathcal{Z}_i(t) = Z_i(t)\,,\ \ \mathcal{F}_e(t) = F_e(t)\,, \tag{13}$$

$$\mathcal{B}_i(t) = [\,B_i(t) + S_i/2\,]_0^{S_i}\ , \tag{14}$$

$$\mathcal{Y}_i(t) = \mathcal{B}_i(t+1) - \mathcal{B}_i(t)\,, \tag{15}$$

$$\mathcal{W}_i(t) = W_i(t) + Y_i(t) - \mathcal{Y}_i(t)\,. \tag{16}$$

In particular, once we solve for the optimal control in the surrogate LQ problem, (15) and (16) tell us what control to use in the actual system. Notice that the actual fast generation control provides the fast generation in the virtual system plus an additional term that keeps the actual storage level process within the hard limit. Note also

$$\mathcal{W}_i(t) \ge W_i(t) - (B_i(t) - S_i/2)_+ - (-B_i(t) - S_i/2)_+\,. \tag{17}$$

Hence the performance parameters $\varepsilon_{\mathcal{F}}$, $\varepsilon_{\mathcal{W}}$ can be estimated from the corresponding ones for the virtual processes.

Now we turn to solving the surrogate LQ problem. First we formulate it in standard state-space form. For simplicity, we will assume $\{Z(t)\}$ is an i.i.d. process (over time)[2] Hence $X(t) := [F(t-1)^T, B(t)^T]^T$ is the state of the system. Also, $U(t) := [Y(t)^T, W(t)^T]^T$ is the control and $R(t) := [X(t)^T, Z(t)^T]^T$ is the observation vector available to the controller. Then

$$X(t+1) = \mathbf{A}X(t) + \mathbf{D}U(t) + \mathbf{E}Z(t), \tag{18}$$

$$R(t) = \mathbf{C}X(t) + \zeta(t)\,, \tag{19}$$

where

$$\mathbf{A} \equiv \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I} \end{bmatrix},\ \ \mathbf{D} \equiv \begin{bmatrix} -\nabla\Delta^{-1} & -\nabla\Delta^{-1} \\ \mathbf{I} & 0 \end{bmatrix},\ \ \mathbf{E} \equiv \begin{bmatrix} \nabla\Delta^{-1} \\ 0 \end{bmatrix}.$$

and $\mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I} \end{bmatrix}$ and $\zeta(t) = \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix} Z(t)$. We are interested in trading off between the performance parameters $\varepsilon_{F_e}, \varepsilon_{W_i}$ and ε_{B_i}'s. Therefore we introduce weights γ_e's , ξ_i's, η_i's and define the Lagrangian

$$\begin{aligned}\mathcal{L}(t) &\equiv \sum_{e=1}^{|E|} \gamma_e \mathbb{E}\{F_e(t)^2\} + \sum_{i=1}^{|V|} \xi_i \mathbb{E}\{B_i(t)^2\} + \sum_{i=1}^{|V|} \eta_i \mathbb{E}\{W_i(t)^2\} \\ &= \mathbb{E}\big\{X(t)^T \mathbf{Q}_1 X(t) + U(t)^T \mathbf{Q}_2 U(t)\big\}\,,\end{aligned} \tag{20}$$

where $\mathbf{Q}_1$, $\mathbf{Q}_2$ are suitable diagonal matrices.

For the sake of deriving optimal filters, we will assume $\mathbb{E}\{Z(t)\} = 0$, and hence all of the virtual processes have 0 mean. If $Z(t)$ has a non-zero mean, this can be subtracted from the system (18): the two systems are equivalent with respect to the minimization of variances, which will be our focus[3] below. From here on we will also assume that $\Sigma_Z \equiv \mathbb{E}[Z(t)^T Z(t)] = \mathbf{I}$, since if not, then we can define $\mathbf{E} = [\nabla\Delta^{-1}\sqrt{\Sigma_Z}^{-1} 0]^T$, where $\sqrt{\Sigma_Z}$ is the symmetrical square root of Σ_Z.

An *admissible control policy* is a mapping $\{R(t), R(t-1), \dots, R(0)\} \mapsto U(t)$. The surrogate LQ problem is defined as the problem of finding the mapping that minimizes the stationary cost $\mathcal{L} \equiv \lim_{t\to\infty} \mathcal{L}(t)$.

Notice that the energy production-minus-consumption $Z(t)$ plays the role both in the evolution equation (18) and the observation (19). The case of correlated noise has been considered and solved for general correlation structure in [9]. Let $\mathbf{G} = \mathbb{E}[\zeta(t)Z(t)^T] = [\mathbf{I}\,0]^T$, $R_1(t) = [Z(t)^T, 0]^T$ and $R_2(t) = [0, B(t)^T]^T$. Adapting the general result in [9] to our special case, we have

[2]If $\{Z(t)\}$ has memory, then one can augment the state space.

[3]We notice in passing that, in the case of general (non-transitive) networks, the appropriate surrogate LQ problem also involves the optimization of means. We defer this aspect to a future publication.

Lemma 3.1: The optimal controller for the system in (18) and (19) and the cost function in (20) is given by

$$U(t) = -(\mathbf{L}R_1(t) + \mathbf{K}^{-1}\mathbf{D}^T\mathbf{S}\mathbf{E}\mathbf{G}^T\mathbf{M}^{-1}R_2(t)), \quad (21)$$

where $\mathbf{S}$ is given by the algebraic Riccati equation

$$\mathbf{S} = \mathbf{A}^T\mathbf{S}\mathbf{A} + \mathbf{Q}_1^T\mathbf{Q}_1 - \mathbf{L}^T\mathbf{K}\mathbf{L}, \quad (22)$$

where $\mathbf{K} = \mathbf{D}^T\mathbf{S}\mathbf{D} + \mathbf{Q}_2^T\mathbf{Q}_2$, $\mathbf{L} = \mathbf{K}^{-1}(\mathbf{D}^T\mathbf{S}\mathbf{A} + \mathbf{Q}_2^T\mathbf{Q}_1)$, and

$$\mathbf{M} = \mathbf{C}\mathbf{J}\mathbf{C}^T + \begin{bmatrix}\mathbf{I} & 0\\ 0 & 0\end{bmatrix}, \quad (23)$$

where $\mathbf{J}$ satisfies the algebraic Riccati equation $\mathbf{J} = \mathbf{A}\mathbf{J}\mathbf{A}^T + \mathbf{E}\mathbf{E}^T - \mathbf{O}\mathbf{M}\mathbf{O}^T$, and $\mathbf{O} = (\mathbf{A}\mathbf{J}\mathbf{C}^T + \mathbf{E}\mathbf{G}^T)\mathbf{M}^{-1}$.

B. Transitive Networks

Lemma 3.1 states that the optimal controller is linear in the storage process and the input noise. However, it is difficult in general to solve analytically the Riccati equations. To gain further insight, we consider the case of transitive networks.

An automorphism of a graph $G = (V, E)$ is a one-to-one mapping $f : V \to V$ such that for any edge $e = (u, v) \in E$, we have $e' = (f(u), f(v)) \in E$. A graph is called *transitive* if for any two vertices v_1 and v_2, there is some automorphism $f : V \to V$ such that $f(v_1) = v_2$. Intuitively, a graph is transitive if it looks the same from the perspective of any of the vertices. Given an electric network, we say the network is transitive if it has a transitive graph structure, every bus has the same associated storage, every line has the same capacity and inductance, and $Z_i(t)$ is i.i.d. across the network. Without loss of generality, we will assume $S_i = S$, $C_e = C$, $B_e = 1$, $\mathbb{E}[Z_i(t)] = \mu, \mathrm{Var}[Z_i(t)] = \sigma^2$). Since the graph is transitive, it is natural to take the cost matrices as $\mathbf{Q}_1 = \mathrm{diag}(\gamma, \ldots, \gamma, \xi, \ldots, \xi)$ and $\mathbf{Q}_2 = \mathrm{diag}(0, \ldots, 0, 1, \ldots, 1)$.

Recall that $\Delta = -\mathbf{V}\alpha^2\mathbf{V}^T$ is the eigenvalue decomposition of Δ. Since $\Delta = -\nabla^T\nabla$, the singular value decomposition of ∇ is given by $\nabla = \mathbf{U}\alpha\mathbf{V}^T$ for some orthogonal matrix $\mathbf{U}$. The basic observation is that, with these choices of $\mathbf{Q}_1$ and $\mathbf{Q}_2$, the Riccati equations diagonalize in the bases given by the columns of $\mathbf{V}$ (for vectors indexed by vertices) and columns of $\mathbf{U}$ (for vectors indexed by edges).

A full justification of the diagonal ansatz amounts to rewriting the Riccati equations in the new basis. For the sake of space we limit ourselves to deriving the optimal diagonal control. We rewrite the linear relation from $X(t)$ to $U(t)$ as

$$Y(t) = \mathbf{H}Z(t) - \mathbf{K}B(t), \quad (24)$$
$$W(t) = \mathbf{P}Z(t) + \mathbf{Q}B(t). \quad (25)$$

Substituting in Eq. (18), we get

$$B(t+1) = (\mathbf{I} - \mathbf{K})B(t) + \mathbf{H}Z(t), \quad (26)$$
$$F(t+1) = \nabla\Delta^{-1}\big\{(\mathbf{I} - \mathbf{H} - \mathbf{P})Z(t) + (\mathbf{K} - \mathbf{Q})B(t)\big\}, \quad (27)$$
$$W(t) = \mathbf{P}Z(t) + \mathbf{Q}B(t). \quad (28)$$

Denoting by $\overline{B}$, $\overline{F}$, $\overline{W}$ the average quantities, it is easy to see that, in a transitive network, we can take $\overline{F} = 0$, $\overline{W} = \mu$ and hence $\overline{B} = 0$. In words, since all nodes are equivalent, there is no average power flow ($\overline{F} = 0$), the average overproduction is dumped locally ($\overline{W} = \mu$), and the average storage level is kept constant ($\overline{B} = 0$).

Hereafter we focus on deviations from the average, and work in the basis in which $\nabla = \mathbf{U}\alpha\mathbf{V}^T$ is diagonal. We will index singular values by $\theta \in \Theta$ hence $\alpha = \mathrm{diag}(\{\alpha(\theta)\}_{\theta\in\Theta})$ (omitting hereafter the singular value $\alpha = 0$ since the relevant quantities have vanishing projection along this direction.) In the examples treated in the next sections, θ will be a Fourier variable. Since the optimal filter is diagonal in this basis, we write $\mathbf{K} = \mathrm{diag}(k(\theta))$, $\mathbf{H} = \mathrm{diag}(h(\theta))$ and $\mathbf{P} = \mathrm{diag}(p(\theta))$, $\mathbf{Q} = \mathrm{diag}(q(\theta))$.

We let $b_\theta(t)$, $z_\theta(t)$, $f_\theta(t)$, $w_\theta(t)$ denote the components of $B(t) - \overline{B}$, $Z(t) - \mu$, $F(t) - \overline{F}$, $W(t) - \overline{W}$ along in the same basis. From Eqs. (26) to (28), we get the scalar equations

$$b_\theta(t) = (1 - k(\theta))b_\theta(t-1) + h(\theta)z_\theta(t), \quad (29)$$
$$f_\theta(t) = -\alpha^{-1}(\theta)\big\{(1 - h(\theta) - p(\theta))z_\theta(t) + (k(\theta) - q(\theta))b_\theta(t-1)\big\}, \quad (30)$$
$$w_\theta(t) = p(\theta)z_\theta(t) + q(\theta)b_\theta(t-1), \quad (31)$$

We will denote by $\sigma_B^2(\theta)$, $\sigma_F^2(\theta)$, $\sigma_W^2(\theta)$ the stationary variances of $b_\theta(t)$, $f_\theta(t)$, $w_\theta(t)$. From the above, we obtain

$$\sigma_B^2(\theta) = \frac{h^2}{1 - (1-k)^2}\,\sigma^2, \quad (32)$$
$$\sigma_F^2(\theta) = \frac{1}{\alpha^2}\left[(1 - h - p)^2 + \frac{h^2(k-q)^2}{1 - (1-k)^2}\right]\sigma^2, \quad (33)$$
$$\sigma_W^2(\theta) = \left[p^2 + \frac{h^2q^2}{1 - (1-k)^2}\right]\sigma^2. \quad (34)$$

(We omit here the argument θ on the right hand side.)

In order to find h, k, p, q, we minimize the Lagrangian (20). Using Parseval's identity, this decomposes over θ, and we can therefore separately minimize for each $\theta \in \Theta$

$$\mathcal{L}(\theta) = \sigma_W(\theta)^2 + \xi\,\sigma_B(\theta)^2 + \gamma\,\sigma_F(\theta)^2. \quad (35)$$

A lengthy but straightforward calculus exercise yields the following expressions.

Theorem 1: Consider a transitive network. The optimal linear control scheme is given, in Fourier domain $\theta \in \Theta$, by

$$p(\theta) = q(\theta) = \xi\,\frac{\sqrt{4\beta(\theta)+1} - 1}{2}, \quad (36)$$
$$h(\theta) = \frac{2\beta(\theta) + 1 - \sqrt{4\beta(\theta)+1}}{2\beta(\theta)}, \quad (37)$$
$$k(\theta) = \frac{\sqrt{4\beta(\theta)+1} - 1}{2\beta(\theta)}, \quad (38)$$

where $\beta(\theta)$ is given by

$$\beta(\theta) = \frac{\gamma}{\xi(\gamma + \alpha^2(\theta))}. \quad (39)$$

It is useful to point out a few analytical properties of these filters: *(i)* $\gamma/[\xi(\gamma + d_{\max})] \le \beta \le 1/\xi$ with $d_{\max}$ the maximum degree in G; *(ii)* $0 \le k \le 1$ is monotone

<table>
<tr><th></th><th>No transmission ($C = 0$)</th><th>No storage ($S = 0$)</th><th>Storage and Transmission</th></tr>
<tr><td>1-D</td><td rowspan="2">$\Theta(\frac{\sigma^2}{S})$ for $\mu S < \sigma^2$
$\sigma \exp\left\{-\frac{\mu S}{\sigma^2}\right\}$ otherwise</td><td>$\Theta(\frac{\sigma^2}{C})$ for $\mu C < \sigma^2$
$\sigma \exp\left\{-\frac{\mu C}{\sigma^2}\right\}$ otherwise</td><td>$\sigma \exp\left\{-\sqrt{\frac{CS}{\sigma^2}}\right\}^\dagger$ for $\mu = \exp\left\{-\omega\left(\sqrt{\frac{CS}{\sigma^2}}\right)\right\}$
$\sigma \exp\left\{-\frac{CS}{\sigma^2}\right\}$ for $\mu = \exp\left\{-o\left(\sqrt{\frac{CS}{\sigma^2}}\right)\right\}$</td></tr>
<tr><td>2-D</td><td>$\sigma \exp\{-\frac{C}{\sigma}\}^\dagger$ for $\mu = \exp\left\{-\omega\left(\frac{C}{\sigma}\right)\right\}$
$\sigma \exp\left\{-\frac{C^2}{\sigma^2}\right\}$ for $\mu = \exp\left\{-o\left(\frac{C}{\sigma}\right)\right\}$</td><td>$\sigma \exp\left\{-\frac{C\max(C,S)}{\sigma^2}\right\}$</td></tr>
</table>

TABLE I

ASYMPTOTICALLY OPTIMAL $\varepsilon_{\mathcal{W}} + \varepsilon_{\mathcal{F}}$ IN 1-D AND 2-D GRIDS.

Logarithmic factors have been neglected (also in the exponent). † indicates the lower bound requires a conjecture in probability theory.

decreasing as a function of β, with $k = 1 - \beta + O(\beta^2)$ as $\beta \to 0$ and $k = 1/\sqrt{\beta} + O(1/\beta)$ as $\beta \to \infty$; (iii) $0 \le h \le 1$ is such that $h + k = 1$. In particular, it is monotone increasing as a function of β, with $h = \beta + O(\beta^2)$ as $\beta \to 0$ and $h = 1 - 1/\sqrt{\beta} + O(1/\beta)$ as $\beta \to \infty$; (iv) $p = q = \xi\beta k$.

Theorem 2: Consider a transitive network, and assume that the optimal LQ control is applied. The variances are given as follows in terms of $k(\theta)$, given in Eq. (37):

$$\frac{\sigma_B^2(\theta)}{\sigma^2} = \frac{(1-k(\theta))^2}{1-(1-k(\theta))^2}, \tag{40}$$

$$\frac{\sigma_F^2(\theta)}{\sigma^2} = \frac{\alpha^2(\theta)}{(\gamma+\alpha^2(\theta))^2}\,\frac{k^2(\theta)}{1-(1-k(\theta))^2}, \tag{41}$$

$$\frac{\sigma_W^2(\theta)}{\sigma^2} = \frac{\gamma^2}{(\gamma+\alpha^2(\theta))^2}\,\frac{k^2(\theta)}{1-(1-k(\theta))^2}. \tag{42}$$

IV. 1-D AND 2-D GRIDS: OVERVIEW OF RESULTS

For the rest of the paper, we focus on two specific network topologies: the infinite one-dimensional grid (line network) and the infinite two-dimensional grid. We will assume Gaussian net supply $Z_i(t) \sim N(\mu, \sigma^2)$. We will focus on the regime when the achieved cost is small. In Section V we will evaluate the performance of the LQ scheme on these topologies. In Section VI. we will derive lower bounds on the performance of any schemes on these topologies to show that the LQ scheme is optimal in the small cost regime. As a result, we characterize explicitly the asymptotic performance in this regime. The results are summarized in Table I.

The parameter μ, the mean of the net supply at each node, can be thought of as a measure of the amount of *over-provisioning*. For the 1-D grid, when the amount of over-provisioning is less than σ^2/C, one can see the dramatic improvement by using storage and transmission resources jointly. When there is only an isolated node with storage, the optimal cost decreases only slowly with the amount of storage S, like $1/S$. Similarly, when there is only transmission but no storage, the optimal cost decreases only slowly with transmission capacity C, like $1/C$. On the other hand, with both storage and transmission, the optimal costs decreases *exponentially* with $\sqrt{CS}$. When there is no storage, the only way to drive the cost significantly down is at the expense of increasing the amount of over-provisioning beyond σ^2/C; the same performance can be achieved with a storage S equalling to this amount of over-provisioning and with the actual amount of over-provisioning exponentially smaller.

The 2-D grid provides significantly better performance than the 1-D grid. For example, the cost exponentially decreases with the transmission capacity C even without over-provisioning and without storage. The increased connectivity in a 2-D grid allows much more spatial averaging of the random net supplies than in the 1-D grid. The reason for this is roughly as follows. Let us focus on the zero over-provisioning case. Consider first a 1-D grid. The aggregate net supply inside a segment of l nodes has variance $l\sigma^2$ and hence the quantity is of the order of $\sqrt{l}\sigma$. This random fluctuation has to be compensated by power delivered from the rest of the grid, but this power can only be delivered through the two links, one at each end of the segment and each of capacity C. Hence, successful compensation requires $l \lesssim C^2/\sigma^2$. One can think of $l^* := C^2/\sigma^2$ as the *spatial scale* over which averaging of the random supplies is taking place. Beyond this spatial scale, the fluctuations will have to be compensated by fast generation. This fluctuation is of the order of $\sqrt{l^*}\sigma/l^* = \sigma/C$ per node. Note that a limit on the spatial scale of averaging translates to a large fast generation cost. In contrast, in the 2-D grid, both the net supply, and the total link capacity connecting a l nodes by l nodes box to the rest of the grid scale up linearly in l. This facilitates averaging over a very large spatial scale l, resulting in a much lower fast generation cost.

There is an interesting parallelism between the results for the 1-D grid with storage and the 2-D grid without storage. If we set $S = C$, the results are in fact identical. One can think of storage as providing an additional dimension for averaging: time (Section VI-B formalizes this). Thus, a 1-D grid with storage is like a 2-D grid without storage.

V. PERFORMANCE OF LQ SCHEME IN GRIDS

In this section we evaluate the performances of the LQ scheme on the 1-D and 2-D grids. Both are examples of transitive graphs and hence we will follow the formulation in Section III-B. The total variance of $B_i(t)$, $F_e(t)$, $W_i(t)$ can be computed from the optimal LQ results, cf. Eqs. (40) to (42) by applying Parseval identity $\sigma^2_{B,F,W} = \int_{[-\pi,\pi]^d} \sigma^2_{B,F,W}(\theta) \mathrm{d}\theta/(2\pi)^d$, with d the grid dimension. Using the fact that B, F, W are Gaussian random variables, and using Eq. (17), we get the following estimates

$$\varepsilon_{\mathcal{F}} \le 2\sigma_F \mathsf{F}\Big(\frac{C}{\sigma_F}\Big), \quad \varepsilon_{\mathcal{W}} \le \sigma_B \mathsf{F}\Big(\frac{S}{2\sigma_B}\Big) + \sigma_W \mathsf{F}\Big(\frac{\mu}{\sigma_W}\Big). \tag{43}$$

Here F is the integral of the tail of a Gaussian random variable $\mathsf{F}(z) \equiv \int_z^{\infty}(x-z)\phi(x)\,\mathrm{d}x = \phi(z) - z\Phi(-z)$, where $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$ the Gaussian density and $\Phi(x) = \int_{-\infty}^{x}\phi(u)\mathrm{d}u$ is the Gaussian distribution. As $z \to \infty$, we have $\mathsf{F}(z) = \frac{1}{z^2\sqrt{2\pi}}\,e^{-z^2/2}\big\{1 + O(1/z)\big\}$.

In order to evaluate performances analytically and to obtain interpretable expressions, we will focus on two specific regimes. In the first one, no storage is available but large transmission capacity exists. In the second, large storage and transmission capacities are available.

A. No storage

In order to recover the performance when there is no storage, we let $\xi \to \infty$, implying $\sigma_B^2 \to 0$ by the definition of cost function (35). In this limit we have $\beta \to 0$, cf. Eq. (39). Using the explicit formulae for the various kernels, cf. Eqs. (36) to (38), we get:

$$p, q = \frac{\gamma}{\gamma + \alpha^2(\theta)} + O(1/\xi)\,,\; h = O(1/\xi)\,,\; k = 1 - O(1/\xi)\,.$$

Substituting in Eqs. (26) to (27) we obtain the following prescription for the controlled variables (in matrix notation)

$$Y(t) = 0 \qquad W(t) = \quad \gamma(-\Delta + \gamma)^{-1}X(t)\,, \tag{44}$$

while the flow and storage satisfy

$$B(t) = 0\,, \quad F(t) = \nabla(-\Delta + \gamma)^{-1}X(t)\,, \tag{45}$$

The interpretation of these equations is quite clear. No storage is retained ($B = 0$) and hence no energy is transferred to storage. The matrix $\gamma(-\Delta + \gamma)^{-1}$ can be interpreted a low-pass filter and hence $\gamma(-\Delta + \gamma)^{-1}X(t)$ is a smoothing of $X(t)$ whereby the smoothing takes place on a length scale $\gamma^{-1/2}$. The wasted energy is obtained by averaging underproduction over regions of this size.

Finally, using Eqs. (41) and (42), we obtain the following results for the variances in Fourier space

$$\frac{\sigma_F(\theta)^2}{\sigma^2} = \frac{\alpha^2(\theta)}{(\gamma + \alpha^2(\theta))^2}\,, \qquad \frac{\sigma_W(\theta)^2}{\sigma^2} = \frac{\gamma^2}{(\gamma + \alpha^2(\theta))^2}\,.$$

1) One-dimensional grid: In this case $\theta \in [-\pi, \pi]$, and $\alpha(\theta)^2 = 2 - 2\cos\theta$ (the Laplacian Δ is digitalized via Fourier transform).

The Parseval integrals can be computed exactly but we shall limit ourself to stating without proof their asymptotic behavior for small γ.

Lemma 5.1: For the one-dimensional grid, in absence of storage, as $\gamma \to 0$, the optimal LQ control yields variances

$$\sigma_F^2 = \sigma^2/4\sqrt{\gamma}\,\Big\{1 + O(\gamma)\Big\}\,, \quad \sigma_W^2 = \sigma^2\sqrt{\gamma}/4\,\Big\{1 + O(\gamma)\Big\}\,.$$

Using these formulae and the equations (43) for the performance parameters, we get the following achievability result.

Theorem 3: For the one-dimensional grid, in absence of storage, the optimal LQ control with Lagrange parameter $\gamma = \mu^2/C^2$ yields, in the limit $\mu/C \to 0$, $\mu C/\sigma^2 \to \infty$:

$$\varepsilon_{\rm tot} = K\,\exp\Big\{-\frac{2\mu C}{\sigma^2}\Big\}\Big[1 + O\Big(\frac{\mu}{C}, \frac{\sigma^2}{\mu C}\Big)\Big]\,, \tag{46}$$

where $K = \sigma^3/\sqrt{32\pi\mu^3 C}$.

The choice of γ given here is dictated by approximately minimizing the cost In words, the cost is exponentially small in the product of the capacity, and overprovisioning μC. This is achieved by averaging over a length scale $\gamma^{-1/2} = C/\mu$ that grows only linearly in C and $1/\mu$.

2) Two-dimensional grid: In this case $\theta = (\theta_1, \theta_2) \in [-\pi, \pi]^2$, and $\alpha(\theta)^2 = 4 - 2\cos\theta_1 - 2\cos\theta_2$. Again, we evaluate Parseval's integral as $\gamma \to 0$, and present the result.

Lemma 5.2: For the two-dimensional grid, in absence of storage, as $\gamma \to 0$, the optimal LQ control yields variances

$$\sigma_F^2 = \frac{\sigma^2}{4\pi}\Big\{\log\Big(\frac{1}{e\gamma}\Big) + O(\gamma)\Big\}\,,\; \sigma_W^2 = \frac{\sigma^2\gamma}{4\pi}\Big\{1 + O(\gamma)\Big\}\,.$$

Using these formulae and the equations (43) for the performance parameters, and approximately optimizing over γ, we obtain the following achievability result.

Theorem 4: For the two-dimensional grid, in absence of storage, the optimal LQ control with lagrange parameter $\gamma = (\mu^2/C^2)\log(C^2/\mu^2 e)$ yields, in the limit $\mu/C \to 0$, $C^2/(\sigma^2\log(C/\mu)) \equiv M \to \infty$:

$$\varepsilon_{\rm tot} = K\exp\Big\{-\frac{2\pi C^2}{\sigma^2\log(C^2/\mu^2 e)}\Big\}\Big[1 + O\Big(\frac{\mu}{C}, \frac{1}{M}\Big)\Big]\,. \tag{47}$$

where $K = \sigma^3\mu\Big(\log(C^2/\mu^2 e)\Big)^{3/2}/(2^{3/2}(2\pi)^2C^3)$ is a polynomial prefactor.

Notice the striking difference with respect to the one-dimensional case, cf. Theorem 3. The cost goes exponentially to 0, but now overprovisioning plays a significantly smaller role. For instance, if we fix the link capacity C to be the same, the exponents in Eq. (46) are matched if $\mu_{\rm 2d} \approx \exp(-\pi C/2\mu_{\rm 1d})\}$, i.e. an exponentially smaller overprovisioning is sufficient.

B. With Storage

In this section we consider the case in which storage is available but overprovisioning μ is small (precise asymptotic conditions will be stated formally below). Within our LQ formulation we want therefore to penalize σ_W much more than σ_B and σ_F. This corresponds to the asymptotics $\gamma \to 0$, $\xi \equiv \gamma/s \to 0$ (the ratio s need not to be fixed). It turns out that the relevant behavior is obtained by considering $\alpha^2 = \Theta(\gamma)$ and hence $\beta \to \infty$. The linear filters are given in this regime by

$$p(\theta) = q(\theta) = (\gamma/\sqrt{s})\,\big(\gamma + \alpha(\theta)^2\big)^{-1/2}\,,$$
$$k(\theta) \approx (1/\sqrt{s})\big(\gamma + \alpha(\theta)^2\big)^{1/2}\,, \quad h(\theta) \approx 1\,.$$

Using these filters we obtain

$$\frac{\sigma_B(\theta)^2}{\sigma^2} \approx \frac{1}{2}\Big(\frac{s}{\gamma + \alpha(\theta)^2}\Big)^{1/2}\,,$$
$$\frac{\sigma_F(\theta)^2}{\sigma^2} \approx \frac{\alpha^2(\theta)}{2\sqrt{s}}\Big(\frac{1}{\gamma + \alpha(\theta)^2}\Big)^{3/2}\,,$$
$$\frac{\sigma_W(\theta)^2}{\sigma^2} \approx \frac{\gamma^2}{2\sqrt{s}}\Big(\frac{1}{\gamma + \alpha(\theta)^2}\Big)^{3/2}\,.$$

1) One-dimensional grid: The variances are obtained by Parseval's identity, integrating $\sigma^2_{B,W,F}(\theta)$ over $\theta \in [-\pi, \pi]$.

Lemma 5.3: Consider a one-dimensional grid, subject to the LQ optimal control. For $\gamma \to 0$ and $\xi = \gamma/s \to 0$

$$\frac{\sigma_B^2}{\sigma^2} = \frac{\sqrt{s}}{4\pi} \log\frac{1}{\gamma} + O(1)\,, \; \frac{\sigma_F^2}{\sigma^2} = \frac{1}{4\pi\sqrt{s}} \log\frac{1}{\gamma} + O(1, s^{-1})\,,$$
$$\frac{\sigma_W^2}{\sigma^2} = \frac{\Omega_1}{2\sqrt{s}}\,\gamma + O(\gamma^2, \gamma^{3/2}/s)\,,$$

where Ω_d is the integral (here $\mathrm{d}^d u \equiv \mathrm{d}u_1 \times \cdots \times \mathrm{d}u_d$)

$$\Omega_d \equiv 1/(2\pi)^d \int_{\mathbb{R}^d} 1/(1 + \|u\|^2)^{3/2}\, \mathrm{d}^d u\,. \tag{48}$$

Using Eqs. (43) to estimate the total cost $\varepsilon_{\rm tot}$ and minimizing it over γ we obtain the following.

Theorem 5: Consider a one-dimensional grid and assume $CS/\sigma^2 \to \infty$. The optimal LQ scheme achieves the following performance:

$$\mu = e^{-\omega\left(\sqrt{\frac{CS}{\sigma^2}}\right)} \Rightarrow \varepsilon_{\rm tot} = \exp\Big\{ -\sqrt{\frac{\pi CS}{2\sigma^2}}\big(1 + o(1)\big)\Big\},$$
$$\mu = e^{-o\left(\sqrt{\frac{CS}{\sigma^2}}\right)} \Rightarrow \varepsilon_{\rm tot} = \exp\Big\{ -\frac{\pi CS(1 + o(1))}{2\sigma^2 \log C/\mu}\Big\},$$

under the further assumption $\sqrt{\pi CS/2\sigma^2} - \log(C/S) \to \infty$ (in the first case) and $\mu^2 \log(C/\mu)/\min(C, S)^2 \to 0$ (in the second). In the first case the claimed behavior is achieved by $s = S^2/4C^2$, and $\gamma = \exp\{-(2\pi CS/\sigma^2)^{1/2}\}$. In the second by letting $s = S^2/4C^2$, and $\gamma = \mu^2 \log(C/\mu)/(\pi\Omega_1 C^2)$.

This theorem points at a striking threshold phenomenon. If overprovisioning is extremely small, or vanishing, then the cost is exponentially small in $\sqrt{CS}$. On the other hand, even a modest overprovisioning changes this behavior leading to a decrease that is exponential in CS (barring exponential factors). Overprovisioning also reduces dramatically the effective averaging length scale $\gamma^{-1/2}$. It also instructive to compare the second case in Theorem 5 with its analogue in the case of no storage, cf. Eq. (46): storage seem to replace overprovisioning.

2) Two-dimensional grid: As done in the previous cases, the variances of B, F, W are obtained by integrating $\sigma^2_{B,F,W}(\theta)$ over $\theta = (\theta_1, \theta_2) \in [-\pi, \pi]^2$.

Lemma 5.4: Consider a two-dimensional grid, subject to the LQ optimal control. For $\gamma \to 0$ and $s = \Theta(1)$, we have

$$\frac{\sigma_B^2}{\sigma^2} = G_B(s) + O(1, \sqrt{\gamma})\,, \qquad \frac{\sigma_F^2}{\sigma^2} = G_F(s) + O(\sqrt{\gamma})\,,$$
$$\frac{\sigma_W^2}{\sigma^2} = \frac{\Omega_2}{2\sqrt{s}}\,\gamma^{3/2} + O(\gamma^2)\,,$$

where Ω_2 is the constant defined as per Eq. (48), and $G_B(s)$, $G_F(s)$ are strictly positive and bounded for s bounded. Further, as $s \to \infty$ $G_B(s) = \mathcal{K}_2\sqrt{s}/2 + O(1)$, $G_F(s) = \mathcal{K}_2/(2\sqrt{s}) + O(1/s)$, where $\mathcal{K}_2 \equiv \int_{[-\pi,\pi]^2} \frac{1}{|\alpha(\theta)|}\, \mathrm{d}\theta$.

Minimizing the total outage over s, γ, we obtain:

Theorem 6: Assume $CS/\sigma^2 \to \infty$ and $C/S = \Theta(1)$. The optimal cost for scheme a memory-one linear scheme on the two-dimensional grid network then behaves as follows

$$\varepsilon_{\rm tot} = \exp\Big\{ -\frac{CS}{2\sigma^2\Gamma(S/C)}\big(1 + o(1)\big)\Big\}\,. \tag{49}$$

Here $u \mapsto \Gamma(u)$ is a function which is strictly positive and bounded for u bounded away from 0 and ∞. In particular, $\Gamma(u) \to \mathcal{K}_2$ as $u \to \infty$, and $\Gamma(u) = \Gamma_0 u + o(u)$ as $u \to 0$ ($\Gamma_0 > 0$).

The claimed behavior is achieved by selecting $s = f(S/C)$, and γ as follows. If $\mu = \exp\{-o(CS/\sigma^2)\}$ then $\gamma = \tilde{f}(S/C)(\mu^2/CS)^{2/3}$. If instead $\mu = \exp\{-\omega(CS/\sigma^2)\}$, then $\gamma = \exp\{-2CS/(3\Gamma(S/C)\sigma^2)\}$, for suitable functions $f, \tilde{f}$ (In the first case, we also assume $\mu/C \to 0$.)

The functions Γ, $f\tilde{f}$ in the last statement can be characterized analytically, but we omit such characterization for the sake of brevity. As seen by comparing with Theorem 5, the greater connectivity implied by a two dimensional grid leads to a faster decay of the cost.

VI. Performance limits

In this section, we prove general lower bounds on the outage $\varepsilon_{\rm tot} = \varepsilon_{\mathcal{W}} + \varepsilon_{\mathcal{F}}$ of *any scheme*, on the 1-D and 2-D grids. Our proofs use cutset type arguments.

A. No storage

1) One-dimensional grid:

Lemma 6.1: Consider a one-dimensional grid without storage. Assume $\mu < C$ and $\sigma < C$. There exists $\kappa_1 > 0$ and $\kappa_2 < \infty$ such that

$$\varepsilon_{\rm tot} \geq \begin{cases} \kappa_1\sigma^2/C & \text{if } \mu < \sigma^2/C\,, \\ \mu\exp\{-\kappa_2\mu C/\sigma^2\} & \text{otherwise.}\end{cases} \tag{50}$$

Proof: Consider a segment of length ℓ. Let E be the event that the segment has net demand at least $3C$. We have

$$\mathbb{P}[\mathsf{E}] \geq \kappa_3 \exp\Big\{ -\frac{(3C + \ell\mu)^2}{\sigma^2\ell}\Big\}\,, \tag{51}$$

for some $\kappa_3 > 0$. If E occurs at some time t, this leads to a shortfall of at least C in the segment of length ℓ. This shortfall contributes either to $\varepsilon_{\mathcal{W}}$ or to $2\varepsilon_{\mathcal{F}}$, yielding

$$2\varepsilon_{\rm tot} \geq \varepsilon_{\mathcal{W}} + 2\varepsilon_{\mathcal{F}} \geq \frac{\kappa_3 C}{\ell}\exp\Big\{ -\frac{(3C + \ell\mu)^2}{\sigma^2\ell}\Big\}\,. \tag{52}$$

Choosing $\ell = \min\big(C/\mu, C^2/\sigma^2\big)$, we obtain the result. ■

Note that the lower bound is tight both for $\mu \geq \sigma^2/C$ (by Theorem 3) and $\mu < \sigma^2/C$ (by a simple generalization of the same theorem that we omit).

2) Two-dimensional grid: We prove a lower bound almost matching the upper bound proved in Theorem 4.

Lemma 6.2: There exists $\kappa < \infty$ such that, for $C \geq \min(\mu, \sigma)$,

$$\varepsilon_{\rm tot} \geq \sigma\exp\Big\{ -\kappa C^2/\sigma^2\Big\}\,.$$

Proof: Follows from a single node cutset bound. ■

We next make a conjecture in probability theory, which, if true, leads to a significantly stronger lower bound for small μ. For any set of vertices $\mathcal{A}$ of the two-dimensional grid, we denote by $\partial\mathcal{A}$ the *boundary* of $\mathcal{A}$, i.e., the set of edges in the grid that have one endpoint in $\mathcal{A}$ and the other in $\mathcal{A}^c$.

Conjecture 6.3: There exists $\delta > 0$ such that the following occurs for all $\ell \in \mathbb{N}$. Let $(X_v)_{v\in\mathcal{S}}$ be a collection of i.i.d. $\mathsf{N}(0,1)$ random variables indexed by $\mathcal{S} = \{1,\dots,\ell\} \times \{1,\dots,\ell\} \subseteq \mathbb{Z}^2$. Then

$$\mathbb{E}\Big[\max_{\substack{\mathcal{A}\subseteq\mathcal{S}\text{ s.t.}\\ |\partial\mathcal{A}|\le 4l}} \sum_{v\in\mathcal{A}} X_v \Big] \ge \delta l \log l \,. \tag{53}$$

Lemma 6.4: Consider the two-dimensional grid without storage, and assume Conjecture 6.3. Then there exists $\kappa < \infty$ such that for any $\mu \le \sigma \exp(-\kappa C/\sigma)$ and $C > \sigma$ we have

$$\varepsilon_{\rm tot} \ge \sigma \, \exp\Big\{ -\kappa C/\sigma \Big\} \,. \tag{54}$$

B. With storage

1) One-dimensional grid: Our approach involves mapping the time evolution of a control scheme in a one-dimensional grid, to a feasible (one-time) flow in a two-dimensional grid. One of the dimensions represents 'space' in the original grid, whereas the other dimension represents time.

Consider the one-dimensional grid, with vertex set $\mathbb{Z}$. We construct a two-dimensional 'spacetime' grid $(\widehat{V}, \widehat{E})$ consisting of copies of each $v \in V$, one for each time $t \in \mathbb{Z}$: define $\widehat{V} \equiv \{(v,t) : v \in \mathbb{Z}, t \in \mathbb{Z}\}$. The edge set $\widehat{E}$ consists of 'space-edges' $E^{\rm sp}$ and 'time-edges' $E^{\rm t}$.

$$\begin{aligned}\widehat{E} &\equiv E^{\rm sp} \cup E^{\rm t}\\ E^{\rm sp} &\equiv \{((v,t),(v+1,t)) : v \in \mathbb{Z}, t \in \mathbb{Z}\}\\ E^{\rm t} &\equiv \{((v,t),(v,t+1)) : v \in \mathbb{Z}, t \in \mathbb{Z}\}\end{aligned}$$

Edges are undirected. Denote by $\widehat{C}_e$ the capacity of $e \in \widehat{E}$. We define $\widehat{C}_e \equiv C$ for $e \in E^{\rm sp}$ and $\widehat{C}_e = S/2$ for $e \in E^{\rm t}$.

Given a control scheme for the 1-D grid with storage, we define the flows in the spacetime grid as

$$\begin{aligned}\widehat{\mathcal{F}}_e &\equiv \mathcal{F}_{(v,v+1)}(t) && \text{for } e = ((v,t),(v+1,t)) \in E^{\rm sp}\\ \widehat{\mathcal{F}}_e &\equiv \mathcal{B}_v(t+1) - S/2 && \text{for } e = ((v,t),(v,t+1)) \in E^{\rm t}\end{aligned}$$

Notice that these flows are not subject to Kirchoff constraints, but the following energy balance equation is satisfied at each node $(v,t) \in \widehat{V}$,

$$\mathcal{Z}_i(t) - \mathcal{W}_i(t) - \mathcal{Y}_i(t) = \sum_{(v',t')\in\partial(v,t)} \widehat{\mathcal{F}}_{(v,t),(v',t')} \tag{55}$$

We use performance parameters as before (this definition applies to finite networks and must be suitably modified for infinite graphs):

$$\begin{aligned}\varepsilon_{\widehat{\mathcal{F}}} &\equiv \frac{1}{|\widehat{E}|}\sum_{e\in\widehat{E}} \mathbb{E}\{(\widehat{\mathcal{F}}_e(t) - \widehat{C}_e)_+ + (\widehat{C}_e - \widehat{\mathcal{F}}_e(t))_+\}\,,\\ \varepsilon_{\mathcal{W}} &\equiv \frac{1}{|\widehat{V}|}\sum_{(i,t)\in\widehat{V}} \mathbb{E}\{\big(W_i(t)\big)_-\}\,.\end{aligned}$$

Notice that $\varepsilon_{\mathcal{W}}$ is unchanged, and $\varepsilon_{\widehat{\mathcal{F}}} = \varepsilon_{\mathcal{F}}$, in our mapping from the 1-D grid with storage to the 2-D spacetime grid.

Our first lemma provides a rigorous lower bound which is almost tight for the case $\mu = e^{-o(\sqrt{CS/\sigma^2})}$ (cf. Theorem 5). It is proved by considering a rectangular region in the spacetime grid of side $l = \max(C/S, 1)$ in space and $T = \max(1, S/C)$ in time.

Lemma 6.5: Suppose $\mu \le \min(C,S)$, $CS/\sigma^2 > \max(\log(C/S),1)$ and $C > \sigma$. There exists $\kappa < \infty$ such that

$$\varepsilon \ge \sigma \exp(-\kappa CS/\sigma^2)\,. \tag{56}$$

Next we provide a sharp lower bound for small μ using Conjecture 6.3. Recall Lemma 6.4 and notice that its proof does not make any use of Kirchoff flow constraints (encoded in Eq. (3)). Thus, the same result holds for a 2-D spacetime grid. We immediately obtain the following result, suggesting that the upper bound in Theorem 5 for small μ is tight.

Theorem 7: There exists $\kappa < \infty$ such that the following occurs if we assume that Conjecture 6.3 is valid. Consider the one-dimensional grid with parameters $C = S > \sigma$, and $\mu \le \exp(-\kappa C/\sigma)$. We have

$$\varepsilon \ge \sigma \exp\Big\{ -\kappa\sqrt{CS/\sigma^2}\Big\}\,. \tag{57}$$

We remark that the requirement $C = S$ can be relaxed if we assume a generalization of Conjecture 6.3 to rectangular regions in the two-dimensional grid.

2) Two-dimensional grid:

Lemma 6.6: There exists a constant $\kappa < \infty$ such that on the two-dimensional grid,

$$\varepsilon \ \ge \sigma \exp\Big\{ -\frac{\kappa C \max(C,S)}{\sigma_i^2}\Big\}\,. \tag{58}$$

The lemma is proved by considering a single node, using a cutset type argument, similar to the proof of Lemma 6.5. It implies that the upper bound in Theorem 6 is tight up to constants in the exponent.

Acknowledgements. This work was partially supported by NSF grants CCF-0743978, CCF-0915145 and CCF-0830796.

REFERENCES

[1] D. MacKay, *Sustainable energy without the hot air*. UIT Cambgridge, 2009.

[2] M. Korpaas, A. Holen, and R. Hildrum, "Operation and sizing of energy storage for wind power plants in a market system," *International Journal of Electrical Power and Energy Systems*, vol. 25, no. 8, pp. 599–606, October 2003.

[3] M. Korpaas, R. Hildrum, and A. Holen, "Optimal operation of hydrogen storage for energy sources with stochastic input," in *IEEE Power Tech Conference Proceedings*, 2003.

[4] J. Paatero and P. Lund, "Effect of energy storage on variations in wind power," *Wind Energy*, vol. 8, no. 4, pp. 421–441, 2005.

[5] P. Brown, J. P. Lopes, and M. Matos, "Optimization of pumped storage capacity in an isolated power system with large renewable penetration," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 523 – 531, May 2008.

[6] B. Nyamdasha, E. Dennya, and M. OMalleyb, "The viability of balancing wind generation with large scale energy storage," *Energy Policy*, vol. 38, no. 11, pp. 7200–7208, November 2010.

[7] H. Su and A. E. Gamal, "Modeling and analysis of the role of fast-response energy storage in the smart grid," in *Proc. Allerton Conference*, 2011.

[8] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Trans. Power Systems*, vol. 24, no. 3, pp. 1290–1300, Aug. 2009.

[9] R. H. Kwong, "On the linear quadratic gaussian problem with correlated noise and its relation to minimum variance control," *SIAM J. Control and Optimization*, 1991.