Satisfying energy demand by harnessing sustainable resources is one of the most important challenges facing our nation and world. In this proposal, I consider the renewable integration problem and propose solutions based on large-scale probabilistic models and optimization-based control. I believe that developing these state-of-the-art machine learning methods to exploit the vast amounts of information now available with the smart grid will be fundamental to integrating our sustainable resources effectively.

The transformation to sustainable resources is already underway; for example, installed wind capacity has increased 4-fold in the past 5 years [3]. However, integrating renewable energy sources into the electrical grid presents fundamental challenges to the way we approach planning and provisioning. In particular, unlike traditional fossil fuels, solar and wind energy is highly variable and non-dispatchable, creating significant problems for operators who must meet demand reliably. Uncertainty also arises due to the difficulty in modeling demand, which although well-studied, remains a open problem. Typically, operators compensate by provisioning slack capacity using forecasting methods. However, the vast majority of existing methods either focus on making point estimates or predict just the marginal distribution for a single location. In reality, these distributions are highly correlated across both time and space; by treating each prediction independently, we lose critical information.

This spatiotemporal forecasting task requires jointly modeling many locations at multiple future time points, resulting in a difficult high-dimensional inference problem beyond the scope of classical methods. However, methods from machine learning excel at modeling high-dimensional distributions using millions of data points; exactly the scenario that now arises with the smart grid. Thus, I propose developing new large-scale probabilistic models for spatiotemporal forecasting and new optimization-based control techniques for using these models in energy planning. By taking a probabilistic approach, we can accurately model real-world correlations which will significantly improve our energy planning capabilities.

First, in my own ongoing work, started in September of this year [4], I consider spatiotemporal forecasting and propose a new model, sparse Gaussian conditional random fields. Concretely, consider the task of forecasting wind power; let $y \in \mathbb{R}^p$ represent wind power forecasts across multiple locations and times and let $x \in \mathbb{R}^n$ represent input features such as past wind power and weather forecasts. The model has a simple log-linear form

$$p(y|x; \Lambda, \Theta) \propto \exp \left\{ -\frac{1}{2} y^T \Lambda y - x^T \Theta y \right\} \qquad (1)$$

where the quadratic term, parameterized by Λ, captures spatiotemporal correlations between wind power forecasts and the linear term, parameterized by Θ, captures the dependence of future wind power on past wind power and weather forecasts. By placing an ℓ_1 penalty on Λ and Θ, the model encourages these parameters to be *sparse*, which allows the model to select the most important correlations amongst the output variables and between the output and input variables. In particular, this model combines the benefits of sparse inverse covariance selection [1] and ℓ_1-regularized regression to jointly model wind forecasts across many locations and times efficiently. In this prior work, I also formulate parameter estimation in this model as a convex optimization problem and develop a novel algorithm to find the

solution. This model achieves state-of-the-art results on forecasting energy demand in the Pennsylvania power grid on a small dataset.

In future research, I plan to extend this model to a wide variety of spatiotemporal forecasting tasks, including forecasting wind and solar power and forecasting the weather systems that affect renewable energy production. The core challenge is in scaling the model to real-world problem sizes since the current optimization algorithm is $O((n+p)^3)$ and cannot be distributed across multiple machines. However, by applying and adapting recent advances in convex optimization, for example Alternating Direction Method of Multipliers techniques [2], I believe I will solve these problems. Developing these new probabilistic models for spatiotemporal forecasting will dramatically improve our ability to quantify uncertainty in many tasks relevant to renewable integration and energy planning. Furthermore, it is important to note that although this model was developed specifically for problems relating to sustainable energy, it also advances the state-of-the-art in machine learning, as it is generally applicable to high-dimensional multivariate regression problems.

Given a high-dimensional probabilistic forecasting model, my next step is to develop control methods that accurately exploit this information to make energy planning decisions. The existing state-of-the-art grid control methods, such as model predictive control, are not well-suited to this task because they are "certainty equivalent," simply taking the mean prediction as input for their optimization. Instead, I will develop an approach based on Monte Carlo methods whereby we sample potential future scenarios from our probabilistic model and plan using multiple runs of model predictive control. This approach naturally handles the spatiotemporal correlations in our forecasting models and adapts accordingly.

By developing state-of-the-art machine learning methods for spatiotemporal forecasting and probabilistic model based control, we can significantly improve the ability of the smart grid to incorporate renewable energy sources. Fundamentally, these methods rely on the availability of large-scale data of the kind that is recently available due to smart meters and sensor networks. By unlocking the power of this data with the methods described in this proposal, we can accelerate the transformation toward sustainable energy with tremendous impact for our country and the entire world.

[1] O. Banerjee, L. El Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9:485–516, 2008.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[3] Various. *Installed Wind Capacity*. U.S. Department of Energy, 2012. Available at: `http://www.windpoweringamerica.gov/wind_installed_capacity.asp`.

[4] M. Wytock and J. Z. Kolter. Sparse Gaussian conditional random fields. In *Workshop on log-linear models (NIPS 2012)*, 2012.