Chapter 6

Conclusion

6.1   Summary of Contributions

This thesis explores the idea of using noise as a tool for understanding gene regulation. Traditionally, cellular noise was considered detrimental because it introduces variability that may prevent a biological circuit from operating in a predictable fashion. As a result, most work focused on understanding how cells mitigate the effect of noise [69, 70, 71, 72]. However, recent work has shown that cells may use noise to their advantage: *B. subtilis* stochastically differentiate into cellular states that allow it to resist conditions where it would otherwise be deprived of nutrients [8]. Viruses use noise to create a bimodal population with two phenotypically distinct sub-populations as a mechanism to evade the host's response [73]. Yeast stress response genes are noisier than general housekeeping genes, allowing a subset of the population to react quickly to environmental stresses [74]. And theoretical work has shown that under certain conditions noisy signals may actually make the response of a gene network highly sensitive [75]. Clearly cells must have mechanisms for dealing with noise and there are many examples where stochastic properties have been exploited.

In this thesis we use noise as a tool for system identification, learning about the time scales, strength, and activity of gene regulation. Cellular noise has not been considered as an engineering tool before, but given its ubiquitous nature in biological systems it would be useful to exploit noise for the purpose of learning about gene regulation. Because the noise occurs naturally it can act as a minimally invasive form of perturbation, and allows for accurate measurements of the cellular conditions that the organism normally experiences.

Here we have shown that noise can be used to measure whether regulation is active, and its temporal properties, using the cross correlation function. We have developed a theoretical framework for calculating the cross correlation between two noisy signals and have used this to predict the properties of experimental systems, both synthetic and natural. We showed that a commonly occurring pattern of regulation, the feed-forward loop, can exhibit a variety of temporal responses that are dependent upon specific model parameters and cellular conditions. In two naturally occurring

feed-forward loops, it was shown that neither is actively regulating its target.

6.2   Future Work

6.2.1   Biological Persistence of Excitation

Engineering work in system identification has proven that to properly identify parameters, a system must be driven by a signal that is sufficiently rich [76]. Without perturbations there are properties of the system's behavior that may go undiscovered; with sufficiently rich inputs the full range of the dynamics are explored. Work on persistence of excitation has shown that for each system parameter to be identified there must be at least one unique frequency in the driving input [77]. White noise, because it excites at all frequencies, has been used as an input [76]. Using noise internal to the system has not been explored in an engineering context, but would be of great utility in understanding biological systems, where significant levels of noise are commonplace.

In this work we do not have direct control over the noisy input to the system and can only measure filtered versions of it. In addition, the frequency of these measurements is limited by practical experimental considerations. It would be interesting to develop a theoretical framework for understanding which model parameters can be identified given realistic sources of biological noise. Given the success of the linearized model, we could start with work on persistence of excitation in linear systems and develop a theory to describe when biological model parameters can be identified.

6.2.2   Network Identifiability

Work in control theory has developed methods for determining when systems are controllable and observable [78]. It would be useful to extend these ideas to a theory of network identifiability. Given a set of measurements and a network structure, an identifiable network is one where the measurements can uniquely determine network parameters. Such a theory would clarify which signals are important to measure and how many measurements are needed.

6.2.3   Monitoring Dynamic Changes

Measurements in this thesis were conducted in conditions where cells had equilibrated to their surrounding environment. It would be interesting to extend these methods to dynamic environments where protein levels are either switching between states, or exhibiting more complex behavior, such as oscillations. In principle, noise-based inference methods can still be applied to dynamically changing systems, though certain approximations about linearized systems may no longer be valid. Experimentally monitoring dynamic behavior may be challenging, too, since the noise-based method relies heavily on averaging across many sets of time-series data.

It may not be necessary to use noise as the perturbing force if the system is already changing dynamically. For example, a system that oscillates may explore enough regions of the state space that noise is not required as a perturbing force.

6.2.4 External Inputs for System Identification

Recent work in microfluidics has made it possible to monitor individual cells while exposing them to time-varying chemical stimuli [79, 80]. This is a more direct way of perturbing the system and measuring its response. Although the cells may not be exposed to purely natural signals, the response properties can be characterized more thoroughly. In addition, applying a known, prescribed input can be an efficient way of learning about the response of a system.

6.2.5 Context-Sensitive Maps of Gene Regulation

Databases of gene regulatory interactions are a good source of information about gene network topology, but lack information about the context in which regulatory elements are active. It will be interesting to explore other examples of regulation to learn when they are active. Although this presents an additional layer of information, the ultimate goal of reducing network structures to the parts that are actually active has the potential to significantly simplify network analysis.