DISSERTATION

ENGINEERING IN PRACTICE:
FROM QUANTITATIVE BIOLOGY MODELING TO ENGINEERING EDUCATION

Submitted by
Lisa Weber
Department of Chemical and Biological Engineering

In partial fulfillment of the requirements
For the Degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Spring 2024

Doctoral Committee:

Advisor: Brian Munsky

Rebecca Atadero
Ashok Prasad
Brad Reisfeld

ABSTRACT

ENGINEERING IN PRACTICE:
FROM QUANTITATIVE BIOLOGY MODELING TO ENGINEERING EDUCATION

In quantitative analyses of biological processes, one may use many different scales of models (e.g., spatial or non-spatial, deterministic or stochastic, time-varying or at steady-state) or many different approaches to match models to experimental data (e.g., model fitting or parameter uncertainty/sloppiness quantification with different experiment designs). These different analyses can lead to surprisingly different results, even when applied to the same data and the same model.

In Chapters 2, a variety of modeling approaches that can be utilized in analyzing biological processes are explained, with examples included of how to mathematically represent a system in order to use these various modeling approaches. Many of these mechanistic modeling approaches are demonstrated in Chapter 3 when we use a simplified gene regulation model to illustrate many of the concerns regarding modeling approach differences; these include ODE analyses of deterministic processes, chemical master equation and finite state projection analyses of heterogeneous processes, and stochastic simulations. For each analysis, we consider a time-dependent input signal (e.g., a kinase nuclear translocation) and several model hypotheses, along with simulated single-cell data, to illustrate different approaches (e.g., deterministic and stochastic) in the identification of mechanisms and parameters of the same model from the same simulated data. We also explore how uncertainty in parameter space varies with respect to the chosen analysis approach or specific experiment design, and conclude with a discussion of how our simulated results relate to the integration of experimental and computational investigations to explore signal-activated gene expression models in yeast [1] and human cells [2].

Different modeling approaches are used in Chapter 4 to build on the work of Scott, *et al.* (2018, 2019) [3, 4] to evaluate different model classes for DNA structural conformation changes, includ-

ing the unwinding/rewinding dynamics of the double-stranded DNA (dsDNA) helical structure and subsequent binding interactions with complementary single-stranded oligonucleotides probes (oligos), in relation to different conditions: temperature, salt concentration, and the level of supercoiling of the DNA molecule. This is done to identify a class of models that best fit the DNA unwinding and subsequent oligo probe binding experimental data as a function of these three conditions. In this work, we demonstrate the use of additional quantitative modeling approaches, including a modified genetic algorithm along with the process of cross-validation and Markov Chain Monte Carlo (MCMC) simulations with the Metropolis-Hastings (MH) algorithm [5] to explore parameter space. We also demonstrate many of the challenges that can be encountered when modeling complex biological phenomena with actual experimental data.

Although much of the work described in Chapters 2 through 4 may appear to be, on the surface, just the use of various computational methods for biological processes to increase understanding of biological mechanisms, much of it also has a separate purpose. The structure of these works and an underlying aim of much of this work, namely Chapters 2 and 3, is to provide guidance with examples to make these computational approaches more accessible to scientists and engineers. Many of these approaches are included in a quantitative biology (UQ-bio) summer school that has been conducted for the last few years as well. Through the process of developing these works and seeking to make quantitative biology more accessible, a related goal manifested to improve the accessibility of engineering education as a whole, which is addressed in Chapter 5, specifically related to diversity, equity, and inclusion (DEI) in undergraduate engineering education. There have been efforts since Fall 2017 to increase the presence of DEI in the undergraduate CBE education using a bottom up approach. To date, various efforts have been incorporated into the first two years of the CBE program. In Chapter 5, these previous efforts, along with lessons learned, are detailed. A substantial, holistic approach to incorporating DEI throughout the CBE curriculum is proposed, based on a review of recent work by other engineering education researchers, to help the CBE department create a more inclusive educational experience for undergraduate students and better enable students to handle the complex challenges they may face in their careers.

ACKNOWLEDGEMENTS

I would foremost like to thank Dr. Brian Munsky and Dr. Rebecca Atadero for your immense support and guidance through these many years, especially your blend of patience, kindness, and intellectual challenge. I can't imagine better mentors and advisors. I would also like to thank Dr. Ashok Prasad for your kindness and mentorship throughout my time at CSU, and thank you for the opportunity to join your summer cancer modeling research. Also, thank you to Dr. Brad Reisfeld and Dr. Ken Reardon for your guidance, advice, and feedback over the years to help me become a more effective, understanding educator. I would also like to acknowledge my collaborators, Will Raymond, Sabrina R. Leslie, Shane Scott, and Cynthia Shaheen. It has been great working with you and gaining insight in a variety of areas. To the past and present Munsky group, including Will Raymond, Michael May, Luis Aguilera, Linda Forero-Quintero, Huy Vo, Zach Fox, Jaron Thompson, Josh Cook, and Eric Ron, thank you for your friendship, feedback, and sharing your knowledge. I would also like to thank Denise Morgan and Claire Lavelle for your help over the years. I would further like to thank Claire Lavelle for her support, rides, and organizing a meal train when I broke my leg this year, as well as thank everyone who volunteered meals. I don't know how my family and I would have managed without your help, and I am immensely grateful. I would also like to acknowledge my funding sources, including the Keck Foundation, NSF (Grant Nos. 1726268, 1725880, 1726088, and 1941870), and the NIH (Subcontract 23X011F from Leidos Biomedical Research, Inc.).

Thank you also to the many friends and family who have supported me over the years. To my brother, I miss you tremendously. The world would be a much better place if you were still here. Most importantly, I thank my husband, Travis, for years of unconditional love, patience, and infinite support. I couldn't have finished this without you and our amazing daughter, Adelaide. Our life together inspires me to keep trying every single day, even when it feels impossible.

DEDICATION

I would like to dedicate this dissertation to my two favorite people,

Travis and Adelaide,

and our wonderful pups, Luka and Freya.

Thank you for helping me find the light when the world is an endless ocean full of dark.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In recent years, massive strides have been made experimentally in the field of biology to both visualize and quantify biological phenomena and biological systems at a single-cell and single-molecule level [3, 4, 6–11]. Along with these advances, however, follows a need to be able to quantitatively analyze the resulting data to understand the biological system and the mechanisms at work within it. Gaining a quantitative understanding of various biological systems, including the dynamics and mechanisms involved, is crucial for numerous applications, among them the development of therapeutics [12], tissue engineering [13–16] and gaining a more comprehensive understanding of the dynamics of cancer on a molecular level [17, 18].

Gaining a quantitative understanding can present a challenge, though, because of the inherent stochastic nature of biological processes, resulting in small copy numbers as well as spatial and temporal fluctuations. These fluctuations lead to both intracellular heterogeneity, or intrinsic noise, and cell-to-cell variability [19, 20], a combination of intrinsic and extrinsic noise, which is noise related to changes within the cellular environment [21]. This variability within and between cells increases the complexity of the system and further complicates the process of mathematically and computationally modeling the system.

To effectively model a biological system, choosing an appropriate scale for the model and modeling approach is of the utmost importance. Mathematical modeling, model fitting, and parameter estimation and uncertainty quantification can be used to model a particular biological system, for example. Additionally, one could choose a deterministic model or a stochastic model, with or without spatial variability, that could be either transient (time varying) or steady state. The selection of a suitable model is not only based upon what knowledge or information one is hoping to gain about the biological system, though, but also upon the nature of the experimental data.

Although many systems have been quantitatively analyzed both computationally and mathematically to date [1, 2, 22–25], many others have only been qualitatively analyzed and are only beginning to be explored quantitatively. The dynamic structural, conformational changes in nucleic acid molecules [3, 4, 26], which can affect their interactions with various other cellular molecules (such as enzymes, proteins, and other molecules involved in or necessary for DNA replication, transcription, and repair) [27–29], are an example of this.

In light of the challenges in understanding and using various modeling approaches for biological systems, including what information can and cannot be elucidated by different methods, how to properly utilize each method, and what are the differences between them, Chapters 2 and 3 in this work walk through a demonstration of this process by first explaining numerous modeling approaches with examples in Chapter 2, and then analyzing the dynamics of a class of simplified multi-state bursting gene regulation models through the use and comparison of multiple mechanistic modeling approaches in Chapter 3. In this process, setting up the stoichiometry matrix and propensity functions for a gene expression model is also described, as is how to formulate the mathematical representation of gene expression in terms of these in state space in order to utilize the many described computational approaches. The quantitative mechanistic modeling approaches described and utilized include: a deterministic model approach with a system of coupled ordinary differential equations (ODEs) to describe bulk gene regulation dynamics and mean gene expression [30–37]; a stochastic simulation approach, which is a dynamic modeling approach that samples from the true distribution of the state space through the use of random variables to evaluate the time evolution of the system [38–50]; and a Chemical Master Equation (CME) based approach which utilizes the finite state projection (FSP) algorithm [51–60] to solve for the flow of probability from state to state within the state space to solve for the distribution within specified error bounds. This work demonstrates model identification and parameter estimation utilizing the methods described above, as well as the associated uncertainty quantification by performing Markov Chain Monte Carlo (MCMC) simulations utilizing the Metropolis-Hastings (MH) algorithm [61, 62]. Additionally, a comparison of the results of these different methods is also included. Extending beyond this, the importance of experiment design is briefly explored by varying an input signal and comparing the differences between the results.

In Chapter 4, to build on the work of Scott, *et al.* (2018, 2019) [3, 4], computational modeling of DNA structural conformation changes, including the unwinding/rewinding dynamics of the double-stranded DNA (dsDNA) helical structure and subsequent binding interactions with complementary single-stranded oligonucleotides probes (oligos), in relation to different conditions are explored. The aim of this work is to identify a class of models that best fit the experimental data of DNA unwinding and oligo probe binding on a particular binding site on the unwound plasmid DNA as a function of three conditions: temperature, salt concentration, and the level of supercoiling of the DNA molecule. To perform this modeling, we use a modified genetic algorithm along with the process of cross-validation to evaluate different model classes. We also use MCMC simulations with the MH algorithm to explore parameter space and attempt to quantify parameter uncertainties. In this work, we demonstrate the use of additional quantitative modeling approaches as well as challenges that can be encountered when modeling complex biological phenomena.

Much of the work described in Chapters 2 through 4 may appear to be, on the surface, just the use of various computational analyses that are performed on biological processes with a goal of simply seeking to increase the understanding of the biological mechanisms at play. However, the structure of these works and the underlying goal of much of this work, Chapters 2 and 3 in particular, is to make these computational approaches more accessible to both scientists and engineers by providing guidance with examples. This work aims to educate, specifically concerning the when, how, and why as related to the applicability and utilization of these different computational approaches. In fact, many of these approaches are included in a quantitative biology (UQ-bio) summer school that has been conducted for the last few years. Through the process of developing these works and seeking to make quantitative biology more accessible, an interest was sparked in the realm of understanding and seeking to improve the accessibility of engineering education as a whole.

In this spirit, the focus of Chapter 5 is diversity, equity, and inclusion (DEI) in undergraduate engineering education, with the primary focal point being the Chemical and Biological Engineering (CBE) Department at Colorado State University (CSU). In this work, an analysis is performed of the current demographics, persistence, and graduation rates for the engineering departments in the Walter Scott, Jr. College of Engineering (COE) at CSU over the last 5-10 years, with a more in depth look at the CBE department. The Covid-19 pandemic resulted in numerous challenges in all departments, particularly as related to enrollment, especially for CBE, and higher education overall is still working to recover. In addition to that, although the department has made notable strides in achieving gender parity in the undergraduate population in the last 3 years, more so when dual degree CBE/BME students are included, not as much progress has been made in regards to historically excluded groups, particularly undergraduate students that are racially and ethnically minoritized. Compared to the State of Colorado statistics for 18-23 year old individuals, the CBE department is considerably less racially and ethnically diverse than the overall Colorado population within the same age demographic. The department also has not made any progress towards achieving gender parity and equivalent female representation in the faculty as compared to the undergraduate population. Since Fall 2017, there have been efforts to increase the presence of DEI in the undergraduate CBE education using a bottom up approach, and to date, various efforts have been incorporated into the first two years of program. In this chapter, these previous efforts are detailed along with lessons learned. Additionally, based on a review of recent work by other engineering education researchers, a substantial, holistic approach to incorporating DEI throughout the CBE curriculum is proposed to help the CBE department to create a more inclusive, supportive educational experience for undergraduate students.

Chapter 2

Quantitative Analysis of Biological Systems[1]

2.1 Introduction

2.1.1 What is quantitative biology?

Quantitative biology (or q-bio) is an interdisciplinary field that integrates biology, computer science, mathematics, and engineering to understand the behavior of biological systems. By employing quantitative methods, researchers have made significant strides in studying vital biological processes, including gene expression [63], signaling pathways [23], viral dynamics [64], microbial community dynamics [25], and more. Quantitative biology is crucial in advancing biomedical research [65], drug development [24], and synthetic biology [66]. The models developed in quantitative biology have far-reaching implications, ranging from designing personalized medical treatments [67] to shaping public policy during pandemics [68].

2.1.2 History of Quantitative Biology

Using quantitative methods to study life sciences is not exactly new. Early examples can be traced back to the 19th century when Pierre-François Verhulst introduced the logistic function to model population growth [69]. In the late 19th century, Cato Guldberg and Peter Waage elucidated the principle for the Law of Mass Action to describe the stoichiometry and rates of elementary chemical reactions [70]. At the beginning of the 20th century, Leonor Michaelis and Maud Menten extended upon these to introduce the model describing the kinetics of substrate and products in enzyme-catalyzed reactions [71]. During the same period, Lotka–Volterra (or predator-prey) equations were introduced to model the evolution of interacting populations of distinct species [72, 73]. During the 50s, Alan Hodgkin and Andrew Huxley introduced a mechanistic model to describe the

---

[1] From a currently in-progress UQ-bio manuscript. Section 2.1.1 has been authored by Luis Aguilera. Section 2.1.2 has been primarily authored by Luis Aguilera, with contributions by me. Sections 2.1.3 and 2.2 have been authored by me.

initiation and propagation of action potentials in neurons [74]. During the ’60s and ’70s, scientists combined experimental observations with statistical models to determine genetic variation and predict quantitative traits in livestock and crops [75]. In 1965, Margaret Dayhoff pioneered the development of computational methods in the study of biological molecules and had been described as the "mother of bioinformatics" [76, 77]. Then, during the 90s, large data sets of DNA, RNA, and protein sequences became available during the genome sequencing era, and the need to store and interpret these data sets gave rise to genomics and structural bioinformatics [77]. By the end of the 90s, important progress was made using mathematical models to understand the evolution of the immune system and HIV in infected patients, paving the way to developing potent antiviral drugs [78]. In parallel, physicists and biologists combined crystallographic data and molecular models to determine the 3D structure of proteins [79]. In the early 2000s, systems biology became a hot topic; this field aims to have a holistic (systems) understanding of the metabolic processes in the cell [80]. At the same time, by combining mechanistic modeling and experimental data, important progress was achieved in understanding the sources of variability (extrinsic and intrinsic noise) in gene expression [81]. More recently, powerful technologies have been introduced to generate massive data sets of biological data, which, combined with more powerful computers, has led to the development of complex models with predictive capabilities and a better understanding of the whole-cell processes [82]. Nowadays, breakthroughs in biology have been achieved by using artificial intelligence and machine learning models. For example, image segmentation has been revolutionized after the introduction of U-net, a deep learning approach capable of automatically segmenting biomedical images [83]. A more recent example is AlphaFold, a deep-learning model that could predict human protein structures with unmatched accuracy [84, 85].

2.1.3 Recent Advances

Significant advances have been made in recent years in experimental techniques to visualize biological systems or quantify biological phenomena at a single-cell or single-molecule level. These include mass-spectrometry based approaches [6], sequencing-based approaches [7], cytometry-based approaches [8], and imaging-based approaches [9, 10], such as single-molecule Fluorescence *in situ* Hybridization (smFISH) [10, 11] and Convex Lens-induced Confinement (CLiC) microscopy [3]. These advances have furthered the collective understanding of cellular processes substantially; however, along with these experimental advances is the challenge of quantitatively analyzing the corresponding data, which can vary dramatically in scale. Although much progress has been made in recent years to apply various mathematical and computational approaches to analyze various biological systems, such as RNA transcription and bursting gene expression [1,2,22], there are numerous biological systems that have been qualitatively analyzed, but are still in the early stages of quantitatively understanding different processes and interactions within that system. An example of this includes the structural dynamics and interactions of nucleic acid molecules. Nucleic acid molecules, such as DNA and RNA, undergo dynamic conformational changes as a result of both external and internal factors, such as temperature, ionic strength, and supercoiling [3,4]. These changes in the structure of the nucleic acid molecules affect interactions of these molecules with enzymes, proteins, and other molecules involved in DNA replication, repair, and transcription [27–29], which are critical for proper cellular function. Quantitatively understanding the dynamics of these biological processes and systems is critical for gaining an understanding of cancer dynamics on a molecular level [17, 18] as well as applications such as the CRISPR-Cas9 gene-editing technique [86,87] and the development of therapeutics [12].

2.2 Quantitative Mechanistic Modeling

Having the ability to simply utilize the numerous possible modeling approaches, however, is inadequate when analyzing biological data. Given the numerous modeling strategies available, it is crucial not only to understand the nature of the data, but also to understand the differences between the information that can be extracted from a given set of experimental data using different modeling approaches. Having a working understanding of the various possible approaches that have been developed to quantitatively analyze data acquired using a particular experimental technique and the differences between modeling approaches is imperative to select an approach (e.g., mathematical modeling, model fitting, or parameter uncertainty quantification) and scale of model (e.g., spatial or non-spatial, deterministic or stochastic, time-varying or steady-state) that is capable of capturing the dynamical behavior of interest, informative, and feasible, given the available data and computational expense. This will enable the selection and utilization of a modeling technique that will maximize the information that can be extracted from the data as well as provide sufficient quantitative information necessary to understand the dynamics of the biological system or process. Often times, these various approaches can lead to significantly different results, even when applied to the same model and same data, because of the stochastic nature of biological processes, as well as the limited data available in many cases because of the difficulty and expense associated with experimentation. If a suitable mathematical or computational modeling approach for a particular dataset is not chosen, one may inadvertently misidentify critical parameters, or inaccurately estimate parameters and their uncertainties, which can result in failing to capture or incorrectly identifying critical system mechanisms and behavior.

In cellular biology, the regulation of genes and the subsequent processes of mRNA transcription and protein translation are inherently stochastic due to the temporal and spatial fluctuations and small copy numbers at the single-cell level, resulting in intracellular heterogeneity and cell-to-cell variability. This can lead to the expression of different phenotypes, even in genetically identical cells within the same chemical environment [19, 20]. The cell-to-cell variability is described as the intrinsic noise of the cell. In contrast, extrinsic noise is related to the cellular environment, such as changes in temperature and nutrients, environmental stresses, and fluctuations in intercellular signals, etc. The stochasticity of gene regulation pathways is the result of a combination of both intrinsic and extrinsic noise [21]. To gain a quantitative understanding of the regulation dynamics in different genes, a system can be analyzed at a *macroscopic* scale using a deterministic approach with differential equations or at a *mesoscopic* scale using a stochastic modeling approach or chemical master equations. The level of complexity and accuracy differs between these modeling approaches, and depending upon the scale of the model and the system behavior of interest, one or more of these approaches may be necessary [88].

2.2.1 Stoichiometry and Propensity Functions

A simple set of biochemical reactions (and their associated stoichiometric coefficients) and their corresponding rates (propensity functions) govern the expression levels over time, or the temporal dynamics, of a gene and its production of mRNA and protein. Once a system has been mathematically characterized using these, the system can be modeled. The mean behavior of the species populations within the system can be evaluated using a deterministic approach, and the temporal fluctuations can be modeled using a stochastic simulations approach. Additionally, the mean of many stochastic simulations can also be compared to the mean behavior identified using the deterministic model of the system [30, 31].

To perform modeling to quantitatively describe gene expression or match models to experimental data, the system must first be described mathematically. This is done by characterizing the system as a set of biochemical reactions. The biochemical reactions related to gene expression are for the transitions from one species (state) to another. These are used to define the (N x M) stoichiometry matrix for the system, with one row for each of the N species and one column for each of the M reactions, which indicates the relative (net) change in the population of each species if that reaction occurs. The possible states of the system can be tracked by assuming that the transition from one state to another is a discrete, random process that depends only upon the *current* state of the system and not previous states or the history of the system, known as Markovian dynamics. We also assume a continuous time evolution, which, when combined with Markovian dynamics, is known as the *well mixed* assumption. The population vector, $\mathbf{x}$, indicates the integer count for every species in the system, where $\mathbf{x} \in \mathbb{Z}^{\mathbf{N}}_{\geq \mathbf{0}}$. The current state of the system, i, is represented by $\mathbf{x}_i = [s_1, s_2, ..., s_N]^T_i$, where s_1 through s_N refer to each of the species in the system. The final state, s_N, is often the species of interest. Any transition between the state, $\mathbf{x}_i$, to the next state, $\mathbf{x}_j$, is denoted as $\mathbf{x}_j = \mathbf{x}_i + \nu_\mu$, where ν_μ represents the stoichiometry associated with the μ^{th} reaction occurring at the i^{th} step.

As an example, consider a bursting gene expression model in which a gene toggles to an active (G_{on}) or inactive (G_{off}) state as a function of time. When in the on state, the gene has two possible transitions that could occur: it could either become inactive (turn off), or it could produce an mRNA molecule through transcription. An mRNA molecule, in turn, could either degrade or produce a protein molecule through translation, which could subsequently degrade. The reactions for this gene expression model are:

(i) $G_{off} \rightarrow G_{on}$ (iii) $G_{on} \rightarrow G_{on} + R$ (v) $R \rightarrow R + P$
(ii) $G_{on} \rightarrow G_{off}$ (iv) $R \rightarrow \phi$ (vi) $P \rightarrow \phi$

For this example, the state vector, $\mathbf{x}$, will have four species: the inactive off gene state (G_{off}), the active on gene state (G_{on}), the number of mRNA in the cell (R), and the number of protein in the cell (P). The population vector for this example is $\mathbf{x} = [G_{off}, G_{on}, R, P]^T$, , where R is the number of mRNA molecules and P is the number of protein molecules. The stoichiometry matrix is:

$$\mathbf{S} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \tag{2.1}$$

where the columns correspond to the stoichiometric coefficients for the reactions (i) through (vi), and the rows correspond to the species in the $\mathbf{x}$ vector. In the matrix, $\mathbf{S}$, the first column is for the reaction (i) $G_{off} \rightarrow G_{on}$. If this reaction were to occur, the population vector $\mathbf{x} = [G_{off}, G_{on}, R, P]^T$ would change by $[-1, 1, 0, 0]^T$.

The rate (or transition probability) associated with each reaction (or state transition) is called its propensity, w, which is a function that is always positive and typically depends upon the current reactant populations and time. The term $w_\mu(\mathbf{x}, t)dt$ is the probability that the μ^{th} reaction initially in state $\mathbf{x}$ will occur in the time-step dt. To determine propensities for individual reactions, the law of mass action can be applied, which states that the rate of a chemical reaction is proportional to the number of unique combinations by which reactants can form products. The propensity functions for each reaction are combined into a column vector when developing a stochastic model.

In the bursting gene expression example, the probability that a single mRNA molecule is transcribed in time $t + dt$ is k_R, where k_R is the transcription rate. The probability that a single mRNA molecule is degraded in time $t + dt$ is $R\gamma_R dt$, where R is the number of mRNA molecules, and γ_R is the degradation rate. The propensity vector for this example, which includes the propensity functions for reactions (i) through (vi), is $\mathbf{w}(\mathbf{x}, t) = [k_{on}G_{off}, k_{off}G_{on}, k_R G_{on}, \gamma_R R, k_P R, \gamma_P P]^T$.

2.2.2 Simulation: Deterministic Modeling Approach

To describe the bulk gene regulation dynamics for the system, a system of coupled ordinary differential equations (ODEs) can be used, where each ODE in the system describes the mean behavior over time for a specific species population. The deterministic description of a system is given by:

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}W(\mathbf{x}, t) \tag{2.2}$$

where $\mathbf{S}$ is the stoichiometry matrix representing the net change of each species after each reaction, and $\mathbf{W}$ is the propensity vector representing the propensity of each reaction in the system. In deterministic modeling and the other modeling approaches that will be discussed, it is assumed that the cell is well-mixed, so molecules are evenly distributed. The ODEs that describe the bursting gene expression example are:

$$\frac{dG_{off}}{dt} = -k_{on}G_{off} + k_{off}G_{on}, \tag{2.3}$$

$$\frac{dG_{on}}{dt} = k_{on}G_{off} - k_{off}G_{on}, \tag{2.4}$$

$$\frac{dR}{dt} = k_R G_{on} - \gamma_R R, \tag{2.5}$$

$$\frac{dP}{dt} = k_P R - \gamma_P P, \tag{2.6}$$

which can be easily solved using an ODE solver. The solutions generated using this approach, while useful for solving for the bulk gene expression dynamics, can be misleading due to the fluctuations exhibited in real systems and since, especially for small volumes like cells, molecule counts aren't continuous. A Python notebook with stoichiometries and propensities and formulating and solving ODEs for gene regulation dynamics can be found at 2021 Undergraduate Qbio Summer School Colab Notebook.

2.2.3 Simulation: Stochastic Simulation Approach

Since the deterministic approach cannot capture stochastic system behavior and model noise, stochastic simulations are often required. Stochastic simulation is a dynamic modeling approach known as a *kinetic Monte Carlo* (kMC) method, which utilizes one or more random variables to sample the time evolution of the state space in a stochastic process. Stochastic simulations generate trajectories that represent how the state of the system and each species' population can change over time, given the stochastic nature of the process and assuming that the population vectors fully describe the system. These trajectories can account for system noise and include the stochastic selection process describing the probability that a reaction will (or will not) occur in time $t + dt$ and, if one does occur, the probability that it will be a particular reaction. If no reaction occurs, nothing in the system will have changed. Since each trajectory is merely a sampling of the probability distribution of possible states at each time t, many SSA trajectories are needed to compute the probability distributions at specific times and to improve the statistical accuracy as related to the mean system behavior [30, 31].

One approach for generating many stochastic trajectories for the state of the system is to use an algorithm such as Gillespie's *direct* stochastic simulation algorithm (SSA) [39, 50, 89]. Two questions must be answered at each step to simulate a stochastic process one reaction at a time. We need to know i) when the next reaction occurs and ii) how the system's state is affected by that reaction. The stochastic simulation algorithm performs the following steps in each iteration:

1. Specify the initial state, $\mathbf{x}(0)$, and the final time, t_{max}.

2. Calculate the propensity functions, $w_1(\mathbf{x})$,...,$w_n(\mathbf{x})$. Then, calculate the sum of the propensities, $w_0(\mathbf{x}) = \sum_{\mu=1}^{M} w_\mu(\mathbf{x})$.

3. Generate the time for the next reaction, τ.

4. Determine which reaction has occurred.

5. Update the current time ($t = t + \tau$) and state of the system ($\mathbf{x} = \mathbf{x} + \sigma_k$).

6. Repeat Steps 2 through 5 while $t < t_{max}$.

To generate the time at which the next reaction will occur (step 3 of the SSA), consider the following: the probability that a reaction will occur in $[t, t + \Delta t)$ is $w(\mathbf{x})\Delta t + \mathcal{O}(\Delta t)^2$, the probability that a reaction will not occur in $[t, t + \Delta t)$ is $1 - w(\mathbf{x})\Delta t + \mathcal{O}(\Delta t)^2$, and the probability that a reaction will not occur in two such time intervals, $[t, t + 2\Delta t)$, is $1 - 2w(\mathbf{x})\Delta t + \mathcal{O}(\Delta t)^2$. Given these, we can determine the probability that no reaction will occur in K time intervals, which is $1 - Kw(\mathbf{x})\Delta t + \mathcal{O}(\Delta t)^2$, assuming K is some positive integer. Letting $\tau = K\Delta t$ and taking the limit as K goes to infinity will yield the probability that no reaction will occur in $[t, t + \tau)$:

$$\lim_{k \to \infty} (1 - w(\mathbf{x})\frac{\tau}{K} + \mathcal{O}(K^{-2}))^K = exp(-w(\mathbf{x})\tau). \tag{2.7}$$

Therefore, the probability that a reaction will occur in $[t, t + \tau)$ is the cumulative probability distribution $F(\tau) = 1 - exp(-w(\mathbf{x})\tau)$, with a probability density function (pdf) of:

$$f(\tau) = \frac{1}{w(\mathbf{x})} exp(-w(\mathbf{x})\tau), \tag{2.8}$$

which is an exponentially distributed (memoryless) random variable. In practice, the time at which the next reaction will occur is a random variable with a distribution given by:

$$P_\tau(t) = |\mathbf{w}(\mathbf{x})|_1 exp(-|\mathbf{w}(\mathbf{x})|_1 t), \tag{2.9}$$

Given this, the time at which the next reaction will occur, τ, can be found by:

$$\tau = \frac{1}{|\mathbf{w}|_1} log \frac{1}{r_1}, \tag{2.10}$$

where $r_1 \in U[0,1]$ and can generated using a uniform random number generator. To evaluate which reaction has occurred at this time (step 4 of the SSA), another uniform random number r_2 is generated, and the smallest k is found, such that:

$$\sum_{\mu=1}^{k-1} w_\mu(\mathbf{x}) \leq r_2|\mathbf{w}|_1 \leq \sum_{\mu=1}^{k} w_\mu(\mathbf{x}), \tag{2.11}$$

Once this is done, the final step (step 5 of the SSA) can be performed, which consists of updating the current time ($t = t+\tau$) and the state of the system ($\mathbf{x} = \mathbf{x}+\sigma_k$). To compute many trajectories, the process will be repeated. One disadvantage of the SSA is that it is quite computationally intensive regarding time and memory. The computational expense increases with the number of reactions, which is why algorithms that can make additional assumptions and reduce computational effort are sometimes utilized.

Examples of more computationally efficient methods that approximate the SSA and generate stochastic trajectories include the *τ-Leaping* Algorithm [90] and the *Chemical Langevin Equation* (CLE) Algorithm [91]. The τ-Leaping Algorithm assumes that there is a time step τ during which the propensities do not change appreciably and, as a result, the number of times the μ^{th} reaction occurs can be approximated by a Poisson-distributed random variable. This method, however, may result in a large decrease in accuracy if the system species have low numbers of molecules and if the propensity functions experience quick, substantial changes. The τ-Leaping Algorithm can be simplified to the CLE Algorithm, which takes advantage of the behavior of *statistically independent* Poisson-distributed random variables with a high density (assuming there is a high number ($\gg 1$) of occurrences of a reaction within the specified time step τ) to be well-approximated by normal distributions with the same mean and variance. Since normal distributions require less computational expenditure to generate, simulations are more efficient if these assumptions are valid. This approach, however, has the same limitations related to accuracy as the *τ-Leaping* Algorithm and is best used with reactions that are sufficiently fast that they can be approximated as continuous-valued Markov processes. Given the limitations of these, though, at least some stochastic simulations using the SSA are often still needed. As the number of stochastic simulations substantially increases, the approximation of the *Chemical Master Equation* (CME) improves. A Python notebook that details the Markov description of gene expression and the Stochastic Simulation Algorithm, as well as demonstrates its use, can be found at 2021 Undergraduate Qbio Summer School Colab Notebook.

2.2.4 Simulation: Chemical Master Equation (CME) Based Approach

The CME (or the *forward Kolmogorov equation*) is a system of linear ordinary differential equations that describe the probability of all possible values for all the molecule species in the system. If solved, the CME has all information to fully describe the evolution of the system. Unfortunately, it is important to consider that for most realistic systems, the state space is infinite, meaning that an infinite number of ODEs are also needed to formulate the CME. Using standard notation, the CME can be derived by knowing (1) the current probability mass for state x_i at time t, $p(\mathbf{x_i}, t)$, and (2) the probability that the system will still be in state x_i at $t + dt$. The CME is:

$$\frac{d}{dt}p(\mathbf{x_i}, t) = -\sum_{\mu=1}^{M} p(\mathbf{x_i}, t)w_\mu(\mathbf{x_i}, t) + \sum_{\mu=1}^{M} p(\mathbf{x_i} - \nu_\mu, \mathbf{t})\mathbf{w}_\mu(\mathbf{x_i} - \nu_\mu, \mathbf{t}) \tag{2.12}$$

To put the CME into matrix form, it is helpful to remember that the CME consists of a linear ODE for each state in the state space, which tracks the flow of probability to and from this state. Consistent with the notation described in 2.2.1, the probability mass vector for the full state space can be defined as $\mathbf{P} \equiv [p(\mathbf{x_1}), p(\mathbf{x_{2}}_{,} ...]$, hence the matrix form of the CME is:

$$\frac{\mathbf{d}}{\mathbf{dt}}\mathbf{P}(\mathbf{t}) = \mathbf{A}\mathbf{P}(\mathbf{t}) \tag{2.13}$$

where the *infinitesimal generator matrix*, $\mathbf{A}$, is equal to:

$$A_{ij} = \left\{ \begin{array}{cl} -\sum_{\mu=1}^{M} w_\mu(\mathbf{x_j}) & \text{for } i = j, \\ w_\mu(\mathbf{x_j}) & \text{for (i,j) such that } \mathbf{x_i} = \mathbf{x_j} + \nu_\mu, \\ 0 & \text{otherwise} \end{array} \right\}. \tag{2.14}$$

2.2.5 Finite State Projection (FSP)

Given the dimensional size of the CME in spite of its linearity, there is rarely an exact or analytical solution that can be found, except for very simple models. Because of this, numerical approaches such as the SSA are used. Another method that can be used to approximate the CME is called the *Finite State Projection* (FSP) Algorithm [51–55]. While the SSA merely samples from a true distribution of the state space, the FSP solves for the distribution within specified error bounds by solving for the flow of probability from state to state within the state space.

To use the FSP algorithm, some necessary additional notation is first introduced, similar to that described in Munsky and Khammash (2008) [53] and Munsky (2012) [51]. If $\mathbf{X} = \{s_1, s_2, ..., s_N\}$, where $s_1, ..., s_N$ are the species in the system, then we can denote a finite subset of $\mathbf{X}$ and its complement, $\mathbf{X_J}$ and $\mathbf{X'_J}$, respectively, that together represent the full state space of the system. Using this notation, $\mathbf{J} = \{j_1, j_2, ..., j_N\}$ is a finite set of indices, and $\mathbf{J'}$ is the complement to the set $\mathbf{J}$, which includes all of the other indices that were not included in $\mathbf{J}$. The full CME can be represented using this notation as:

$$\frac{d}{dt}\begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{JJ} & \mathbf{A}_{JJ'} \\ \mathbf{A}_{J'J} & \mathbf{A}_{J'J'} \end{bmatrix} \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} \tag{2.15}$$

Furthermore, let $\mathbf{A_{IJ}}$ denote a submatrix of $\mathbf{A}$ with rows chosen according to $\mathbf{I}$ and columns chosen according to $\mathbf{J}$. For example, if $\mathbf{A}$ is given as:

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix},$$

and $\mathbf{I}$ and $\mathbf{J}$ are defined as $\{1, 2\}$ and $\{3, 1\}$, respectively, then $\mathbf{A_{IJ}}$ is defined as:

$$\mathbf{A_{IJ}} = \begin{bmatrix} c & a \\ f & d \end{bmatrix},$$

since $[A_{13}, A_{11}] = [c, a]$ and $[A_{23}, A_{21}] = [f, d]$. For convenience, let $\mathbf{A_J} = \mathbf{A_{JJ}}$. If the full CME, shown in Eq. 2.2.4, describes an infinite state Markov process, $\mathcal{M}$, with an initial distribution of $\mathbf{P}(0)$, then the master equation for a finite state Markov process, $\mathcal{M}_\mathbf{J}$, is:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ g(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A_J} & \mathbf{0} \\ -\mathbf{1^T A_J} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ g(t) \end{bmatrix} \tag{2.16}$$

where $g(t)$ consists of one or more aggregate absorbing states for any probability mass that leaves $\mathbf{X_J}$. In this formulation of the master equation, once mass leaves $\mathbf{X_J}$, it cannot return and remains in the absorbing state(s). The initial distribution for this is given by:

$$\begin{bmatrix} \mathbf{P}_J^{FSP}(0) \\ g(0) \end{bmatrix} \begin{bmatrix} \mathbf{P}_J(0) \\ 1 - \sum \mathbf{P}_J(0) \end{bmatrix} \tag{2.17}$$

An important aspect of the FSP formulation of the CME, Eq. 2.2.5 is related to the absorbing state or states, $g(t)$, which contain the *exact probability mass* that has left $\mathbf{X_J}$ at any time, $\tau \in [0, t]$, since probability can only leave $\mathbf{X_J}$, but cannot return in this formulation of the CME. Directly related to this, if probability mass in $\mathbf{X_J}$ has never left $\mathbf{X_J}$ at any time, $\tau \in [0, t]$, then the elements of $\mathbf{P}_J^{FSP}(t)$ must be the exact joint probabilities that the system is in $\mathbf{X_J}$ at time t and has never left $\mathbf{X_J}$. From this, we can posit that $\mathbf{P}_J(t) \geq \mathbf{P}_J^{FSP}(t) \geq 0$ for any index set J since $\mathbf{P}_J^{FSP}(t)$ is a more restrictive joint distribution than $\mathbf{P}_J(t)$. The exact residual of the $\mathbf{P}_J^{FSP}(t)$ as compared to $\mathbf{P}_J(t)$ can be computed as [51]:

$$\begin{aligned}
\left|\left|\begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}'_J(t) \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ 0 \end{bmatrix}\right|\right|_1 &= |\mathbf{P}_J(t) - \mathbf{P}_J^{FSP}(t)|_1 + |\mathbf{P}'_J(t)|_1, \\
&= |\mathbf{P}_J(t)|_1 - |\mathbf{P}_J^{FSP}(t)|_1 + |\mathbf{P}'_J(t)|_1, \\
&= 1 - |\mathbf{P}_J^{FSP}(t)|_1, \\
&= g(t)
\end{aligned} \tag{2.18}$$

Similarly to the SSA approach discussed in Sec. 2.2.3, the stoichiometry matrix and propensity functions are needed for all species as well as the initial probability density vector, $\mathbf{P}(0)$, and the final time, t_f. An input that isn't needed in the SSA, but is required for the FSP is the total amount of acceptable error, $\epsilon > 0$.

With those inputs, the Finite state projection (FSP) algorithm is as follows [51, 58]:

0. Specify the initial finite set of states, $\mathbf{X_J}$, for the FSP.
   Initialize a counter, $i = 0$.

1. Use the propensity functions and stoichiometry to form $\mathbf{A}_{J_i}$.
   Compute $g(t_f)$ by solving Eq. 2.2.5.

2. If $g(t_f) \leq \epsilon$, **Stop**.
   $\mathbf{P}^{FSP}(t)$ approximates $\mathbf{P}(t_f)$ to within a total error of ϵ.

3. Add more states to find $\mathbf{X_{J_{i+1}}}$.
   Increment i and return to Step 1.

There are numerous ways to accomplish these steps, however, as outlined in Munsky (2012) [51].

2.2.6 Summary

Together, the mechanistic modeling approaches described in detail here can be used to analyze both deterministic and stochastic aspects of a biological system. This can be incredibly useful for better understanding the behavior and physical mechanisms of biological phenomena. In the next chapter, these modeling approaches will be utilized to analyze a gene regulatory model. This chapter will further demonstrate the differences between these modeling approaches, what information can be obtained from them, and challenges in using them and interpreting the resulting data.

Chapter 3

Identification of Gene Regulation Models from Single-Cell Data[2]

3.1 Introduction

Recent years have led to rapid advances in the capabilities to measure gene regulatory phenomena at the level of single-cells and in fluctuating conditions. These techniques include image-based strategies, where fluorescence markers are used to highlight specific DNA, RNA, or protein molecules of interest [9,10]; cytometry-based approaches, where cellular properties, especially size or fluorescence, are rapidly measured one cell at a time [8]; mass-spectrometry based approaches, where cellular proteins are digested into peptide fragments whose distinct mass or charge spectra can be analyzed quantitatively [6]; and sequencing approaches, where the abundances of specific RNA or DNA molecules are quantified on a single-cell basis [7]. In this article, we focus on the quantitative analysis of one specific technique, known as single-molecule Fluorescence *in situ* Hybridization (smFISH) [10, 11], which provides precise and reproducible measurements of the number and locations of individual RNA molecules within single cells. In the technique of smFISH, researchers tile individual molecules of endogenous RNA with many complementary DNA oligomers. Each RNA-binding oligomer contains one or more fluorescent dyes, such that the entire multi-labeled RNA molecule appears as a bright diffraction limited spot under the fluorescence microscope. Counting the integer numbers of these spots for many single cells in large populations at many times and in the same conditions, one can capture the temporal distributions of gene regu-

---

[2] This is the Accepted Manuscript version of an article accepted for publication in Physical Biology. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The Version of Record is available online at [10.1088/1478-3975/aabc31] [92]. A substantially reduced version also appeared as an extended exercise in the textbook *Quantitative Biology: Theory, Computational Methods and Examples of Models*, edited by Brian Munsky, Lev S. Tsimring, and William S. Hlavacek [93]. This work has been primarily authored by me. The GUI presented in Section 3.5 was developed by William Raymond. In addition, Section 3.4.6 was authored by Brian Munsky.

lation activity for those conditions. By using microfluidics or by adding or removing stimuli, one can observe how single-cell statistics are affected by transitions from one environmental condition to another [1, 2, 22].

Utilizing experiments such as smFISH, many researchers have quantified time-varying and bursty gene expression at the single-cell level, and many models have been developed to capture these dynamics (e.g., see recent reviews at [56, 57, 94]). In this article, we define a small class of single-cell models which have previously been shown to capture and predict smFISH data [1, 2, 22], and we employ several computational analyses in attempts to identify these models and their parameters from various sets of simulated data.

3.1.1 A Standard Computational Toolbox for Quantitative Biology

In the following paragraphs, we introduce a handful of computational tools that are frequently used in the analysis and fitting of quantitative gene regulation models.

Ordinary Differential Equations. One of the most common approaches to describe the dynamic behavior of gene regulation is to use coupled sets of Ordinary Differential Equations (ODEs). For example, the ODE related to mRNA transcription is a mathematical equation which describes the rate of change in the number of mRNA over time, taking into consideration both production and degradation of mRNA molecules. For most biological processes of interest, analytical solutions to the corresponding ODEs may not be possible and numerical techniques may be necessary. In this article, we will deal with reaction rates that change over time, for which numerical integration is required.

Parameter Fitting and the Metropolis-Hastings Algorithm. A common task in quantitative biology is to fit parameters of a model to match experimental data. Due to the stochastic nature of biological systems, in addition to the fact that experimental data is noisy and limited in quantity, exact estimation of parameters is impossible. Having a quantitative understanding of parameter uncertainties in light of existing data is crucial for predictive modeling. To accomplish this goal, Markov Chain Monte Carlo (MCMC) analyses, such as the Metropolis-Hastings (MH) algorithm [61], are frequently used to provide unbiased samples of parameter space and to estimate parameter uncertainties. In particular, the MH algorithm samples parameter space by adding random perturbations to parameters and then calculating the likelihood of the data, given the new parameter set. The new parameter set is then discarded unless the likelihood exceeds an appropriately chosen (albeit) random acceptance threshold. Like any parameter fitting routine, the MH algorithm has several tunable parameters. The *MH chain* is the sequence of parameter sets that have been accepted during the random walk through parameter space. The *proposal distribution* is a probabilistic rule by which a new parameter set is generated from the old. The *burn-in* period is the number of MH samples that are ignored to allow the MH chain to relax away from the (potentially poor) initial parameter guess. The *chain length* is the total number of MH samples that are recorded following the burn-in period. The *thin rate* is the number of accepted samples that are ignored between those that are recorded (i.e., every 10^{th} sample). See [95] for a more in depth discussion of Monte Carlo analyses (Chapters 10 and 16), parameter estimation (Chapter 13), and sensitivity analyses (Chapter 14) as applied to models in quantitative biology.

The Stochastic Simulation Algorithm. Unlike most dynamical problems in science and engineering, gene regulation is inherently subject to discrete stochastic behavior arising from the single-molecule nature of genes within individual cells. The standard approach to analyze such discrete stochastic processes is to apply the kinetic Monte Carlo algorithm known as (Gillespie's) *Stochastic Simulation Algorithm* (SSA) [38, 50], which uses a combination of stochastic reaction rates (i.e., propensity functions) and reaction rules (i.e., stoichiometry vectors) to simulate the stochastic trajectories. This algorithm updates the state of the system after each random reaction and assumes exponentially-distributed waiting times between every event. In this article, we apply the SSA to simulate many possible stochastic trajectories for the switching of gene activity states coupled with mRNA production and degradation.

The Finite State Projection Method. In addition to using the SSA to generate stochastic trajectories for the gene regulation processes, we also explore direct computations of the underlying probability distributions for these processes. These distributions evolve according to the infinite dimensional linear equation known as the Chemical Master Equation (CME) [60, 96]. Because the CME has no exact solution for most models, we employ the finite state projection (FSP) method to approximate the CME solution and to estimate the full joint probability distributions for gene switching behavior and transcript accumulation [52]. Several recent reviews of the FSP approach and its applications can be found at [55, 57, 58, 97]. A key advantage of the CME and FSP formulation is that they provide systematic approaches to compute and compare the likelihood of data arising from multiple models [1, 2, 55]. With this in mind, we use the Metropolis-Hastings algorithm in conjunction with the FSP analysis to attain more precise constraints on the parameters when fit to simulated data [22].

Experiment Design. The successful identification of a model from data requires not just good data and appropriate modeling tools, but the experiment itself also needs to be conducted using a well-designed and informative condition [58, 98–103]. With this in mind, we will explore the importance of experiment design, and we will show that different experiments (e.g., different input signals and different measurement times) and different analyses (e.g., ODEs or the FSP) yield different results for the constraint of system parameters.

In the following section, we introduce a specific gene regulation model and its underlying assumptions. Then, in Section 3.3, we discuss the simulated data from which we seek to identify our model parameters. In Section 4.2, we demonstrate the implementation of the above tools to this model and different sets of simulated data. Finally, in Section 3.5, we provide a brief summary of our results and conclusions, and we introduce a user-friendly PYTHON-based toolbox with a graphical-user-interface to allow readers to reproduce all of the described analyses. All analyses discussed in this article have been implemented in MATLAB and PYTHON codes, and the codes are available for download at: https://github.com/MunskyGroup/WeberPB

3.2 Gene Regulation Model Description

We begin by specifying a simple semi-mechanistic model for gene regulation. Figure 3.1 illustrates this model, which is a generalization of the standard bursting gene expression model discussed in [56,94,104]. The model consists of a single gene with three possible states $\{S_1, S_2$, and $S_3\}$, each of which represents a different configuration of transcription factor binding or chromatin conformation. Similar three-state models have been used previously in the context of time-varying inputs to capture and predict signal-activated single-cell gene expression in yeast [1,22] and human [2] cells. The model assumptions are as described with additional detail in Subsection 3.2.1.

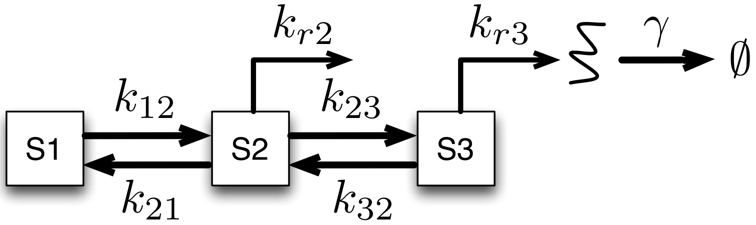


Figure 3.1: Schematic of a three-state gene regulation model. State 1 is transcriptionally inactive, whereas states 2 and 3 are transcriptionally active with transcription rates k_{r2} and k_{r3}, respectively. Parameters k_{ij} denote the (potentially time-varying) transition rates between states S_i and S_j.

3.2.1 Assumptions of the Three-State Bursting Gene Expression Model

First, we define the conformational states $\{S_1, S_2$, and $S_3\}$ of the gene and the production and degradation of mRNA. In this model, S_1 is a transcriptionally inactive ('OFF') state, whereas S_2 and S_3 are transcriptionally active ('ON') states. The transcription rates for S_2 and S_3 are k_{r2} and k_{r3}, respectively. Degradation of mRNA species, R, is assumed to be a first order process with rate γ.

Next, we define the transition dynamics that allow the gene to switch among these states. In the three-state model, there are four state transition events: $S_i \to S_j$, each with a corresponding transition rate of $k_{ij}(t)$. Three of the four transition rates are assumed to be constant with respect to time and the activating kinase: $k_{ij}(t) = k_{ij}(0)$. The gene is activated by a time-varying kinase, $Y_i(t)$, which acts as an input to the system. In practice, such signals can be controlled experimentally through modulation of external stimuli [1,2,22,105–107]. For illustrative purposes, we initially assume a sinusoidal input signal given by:

$$Y_1(t) = \left\{ \begin{array}{ll} 0 & \text{for } (t \leq 5)(t \geq 70) \\ 1 - \cos\left(\frac{2\pi t}{30}\right) & \text{for } t \in (5, 70) \end{array} \right\}, \tag{3.1}$$

where the input signal (in arbitrary units of concentration) begins at time t = 5 min and ends at time t = 70 min. We assume exactly one of the four transition rates is linearly dependent upon the activating kinase as follows [1]:

$$k_{ij}(t) = \max\left\{0, k_{ij}(0) + \beta Y_i(t)\right\}. \tag{3.2}$$

For the combination $i < j$ (i.e., when the reaction leads to a state with a higher index), the value of $k_{ij}(0)$ is negative and β is positive. Conversely, when $i > j$ (i.e., when the reaction leads to a state with a lower index), the value of $k_{ij}(0)$ is positive and β is negative. This choice gives rise to a thresholded activation response in which $k_{ij}(t)$ is zero until the kinase exceeds or drops below a specific level. This definition leads to activation, either through direct activation when $i < j$ or indirect loss of repression when $i > j$.

One of our goals in this study is to determine if it is possible to identify the mechanism of action, i.e., to determine which k_{ij} depends upon the time-varying kinase signal. We define the model number, N_m, as the integer that specifies which mechanism depends upon the activating kinase. The model numbers and the corresponding input-dependent transition rates are as follows:

Model 1: $k_{12}(t)$; Model 2: $k_{23}(t)$; Model 3: $k_{21}(t)$; Model 4: $k_{32}(t)$.

Now that we have specified the assumptions needed to define the gene states, mRNA levels, and the associated dynamics, we can write a set of four ODEs to describe this model:

$$\frac{dS_1}{dt} = -k_{12}S_1 + k_{21}S_2, \tag{3.3}$$

$$\frac{dS_2}{dt} = k_{12}S_1 - (k_{21} + k_{23})S_2 + k_{32}S_3, \tag{3.4}$$

$$\frac{dS_3}{dt} = k_{23}S_2 - k_{32}S_3, \tag{3.5}$$

$$\frac{dR}{dt} = k_{r2}S_2 + k_{r3}S_3 - \gamma R. \tag{3.6}$$

These equations are easily integrated in MATLAB using the stiff ODE solver, ODE15S.

Finally, for parameter estimation, we assume that all rates have an independent prior given by a lognormal distribution with a logarithmic mean of $\mathbb{E}\{\log_{10}(\lambda)\} = 0$ and a wide logarithmic variance of $\mathbb{E}\{(\log_{10}(\lambda))^2\} = 4$. Having defined the reactions and priors on all parameters, in the next section we will generate some simulated data and attempt to fit our model, after which we will explore uncertainty in parameter space, and make model predictions in light of these datasets.

3.3 Single-cell Data Collection

The technique of smFISH [10,11] makes it possible to measure the number of mRNA molecules within hundreds or thousands of individual cells at specific instances in time during a dynamic response from one condition to another. For example, in [1, 22], the authors measured and found a quantitative FSP-based model to predict the statistics of *STL1*, *CTT1*, and *HSP12* mRNA expression in yeast cells every two to five minutes following osmotic shock and temporal activation of the HOG1p kinase. Similarly, in [2], the authors measured time-varying p-ERK kinase and its effect on *c-Fos* expression in human-derived U20S cells.

To generate simulated data of this form, we assume a “true" model, which corresponds to Model 2, and for which the parameters are presented in vector form as:

$$\begin{aligned}\mathbf{\Lambda}_{\text{True}} &= [k_{12},\ k_{23},\ k_{21},\ k_{32},\ k_{r2},\ k_{r3},\ \beta,\ \gamma] \\ &= [0.20,\ -0.076,\ 0.050,\ 0.050,\ 16.0,\ 5.0,\ 0.10,\ 1.0]. \end{aligned} \tag{3.7}$$

Throughout this article, all rate parameters will be assumed to have units of (minutes)$^{-1}$, and β has units of (minutes)$^{-1}$(concentration)$^{-1}$.

At the initial time ($t = 0$), we assume that all cells are in state S_1 and that there are no mRNA ($R = 0$), meaning that all cells are in a transcriptionally inactive state. This initial condition was chosen arbitrarily, but matches to observed dynamics for many genes (e.g., *STL1*, *CTT1*, *HSP12*, *c-Fos*), which are transcriptionally silent in the absence of their activating kinase signal [1,2]. We use the FSP approach [52,55,57,58,97] with the true model and known initial condition to compute the time-varying mRNA probability distribution at each combination of time and input condition. We then use the FSP solution to generate histograms of 100 exact independent random samples (i.e., one data point for each simulated cell) at ten equally spaced times between zero and 100 minutes. In the following sections, we will use various computational approaches to infer model parameters from these time-varying histograms.

3.4 Results

The goal of our study is to determine if simulated data is sufficient to identify the model and its parameters, and if the identified parameters can be used to predict the system’s response under new experimental conditions.

3.4.1 Simulation of ODE Models

We first use a deterministic approach and solve the simple system of ODEs listed in Eq. 3.3 - 3.6 to capture the bulk gene regulation dynamics. For a given model and parameter set, these ODEs generate the mean gene expression as a function of time. Figure 3.2 shows the predicted mean gene expression as a function of time for Model 2 (i.e., where k_{23} is time-dependent), assuming the input signal in Eq. 3.1, and two different arbitrary parameter sets (blue and magenta) as given in the caption.

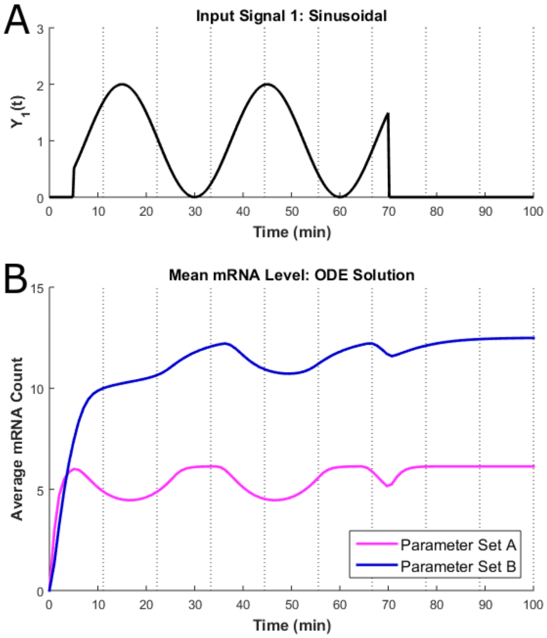


Figure 3.2: (A) Input signal, $Y_1(t)$ in arbitrary units of concentration (AUC) versus time, as defined in Eq. 3.1. (B) Deterministic ODE solution for the mean gene expression as a function of time for the three-state gene regulation model shown in Fig. 3.1, as outlined in Section 3.4.1. Analyses are shown for Model 2 (k_{23} is time dependent), and the two illustrated parameter sets are $\mathbf{\Lambda_A} = [3,\ -0.4,\ 2.5,\ 1,\ 9,\ 2,\ 1,\ 0.8]$ and $\mathbf{\Lambda_B} = [1,\ -0.6,\ 1,\ 0.2,\ 5,\ 2,\ 1,\ 0.2]$.

Examining how different parameters affect system behavior reveals qualitative insight about how rate parameters control the underlying process dynamics. For example, Fig. 3.2 illustrates that faster rates for transcription and degradation (k_{r2} and γ) enable gene expression to track the input signal more closely (Fig. 3.2, magenta line), whereas slower rates result in greater averaging and lag behind the input signal (Fig. 3.2, blue line). Biologically, evolution could tune the rates of these reactions to allow cells to respond more quickly to important fluctuations (e.g., sustained stresses) while rejecting faster environmental disturbances (e.g., short-lived transient fluctuations). In addition, such dependence of response dynamics on parameters makes it feasible to perform an inverse analysis (i.e., to estimate parameters from quantitative observations of response dynamics), as discussed in the following subsection.

3.4.2 Maximum Likelihood Estimation for ODE Models

Next, we use the ODE model defined in Eq. 3.3-3.6 and search parameter space to find a parameter set that maximizes the likelihood of the simulated dataset. As described in Section 3.3, the data were generated at ten equally-spaced time points between zero and 100 minutes, assuming the known input signal, $Y_1(t)$, as given in Eq. 3.1. Prior to fitting with the ODE model, we computed the mean, standard deviation, and standard error of the mean (see Fig. 3.3).

Before we can fit the model to the data, we must first compute the likelihood of the data, given the model's parameters. Consider a set of experimental data, including N_i cells at each time t_i. Let $\tilde{R}_{ij}$ denote the number of mRNA in the j^{th} cell at the i^{th} time point. The measured sample mean at each time is given by:

$$\mu_S(t_i) = \frac{1}{N_i}\sum_{j=1}^{N_i} \tilde{R}_{ij}. \tag{3.8}$$

To fit the ODE model to the measured sample mean, one must quantify how well one would expect the sample mean to match to the predictions made by the ODE analysis. This question is easily answered by invoking the central limit theorem (CLT) [108], which applies provided that the true distribution has finite variance. According to the CLT, when many independent and identically distributed measurements are taken from the same distribution, the average of those measurements will tend to a Gaussian random variable with mean equal to the true mean and variance equal to the true variance divided by the number of samples (or squared standard error). In turn, the squared standard error can be estimated from the measured histograms according to:

$$\sigma^2_{SEM}(t_i) = \frac{1}{N_i(N_i-1)}\sum_{j=1}^{N_i}\left(\tilde{R}_{ij} - \mu_S(t_i)\right)^2. \tag{3.9}$$

Provided that the underlying distribution has a finite true variance and if enough sample measurements are collected, the likelihood to observe the measured sample mean can be estimated using the normal distribution:

$$Prob(\mu_S(t_i) \mid \mathbf{\Lambda}) = \frac{1}{\sqrt{2\pi\sigma^2_{SEM}(t_i)}} \exp\left(-\frac{(\mu_S(t_i) - R(t_i, \mathbf{\Lambda}))^2}{2\sigma^2_{SEM}(t_i)}\right). \tag{3.10}$$

Our assumed lognormal prior distribution ($\mathbb{E}\{log_{10}(\lambda)\} = 0$, $\mathbb{E}\{(log_{10}(\lambda))^2\} = 4$) for each parameter can be written as:

$$Prob(\lambda_k) = A \exp\left(-\frac{\log_{10}(\lambda_k)^2}{8}\right), \tag{3.11}$$

where A is a normalization constant. The probability of a parameter set $\mathbf{\Lambda}$, given the data, can be written according to Bayes' rule, (i.e., $P(\mathbf{\Lambda}|\mathbf{D}) = P(\mathbf{D}|\mathbf{\Lambda})P(\mathbf{\Lambda})/P(\mathbf{D})$, where $\mathbf{\Lambda}$ denotes the model and $\mathbf{D}$ denotes the data) as:

$$Prob(\mathbf{\Lambda}|\mathbf{D}) = \frac{1}{P(\mathbf{D})}\left(\prod_{i=1}^{N_t} Prob(\mu_S(t_i) \mid \mathbf{\Lambda}) \cdot \prod_{k=1}^{8} Prob(\lambda_k)\right). \tag{3.12}$$

Here, N_t is the number of time points, and $P(\mathbf{D})$ is a normalization constant. Taking the logarithm of $Prob(\mathbf{\Lambda}|\mathbf{D})$ and substituting in the expressions for $Prob(\mu_S(t_i) \mid \mathbf{\Lambda})$ and $Prob(\mathbf{\Lambda}|\mathbf{D})$ above yields:

$$\log(Prob(\mathbf{\Lambda}|\mathbf{D})) = C - \sum_{i=1}^{N_t} \frac{(\mu_S(t_i) - R(t_i, \mathbf{\Lambda}))^2}{2\sigma^2_{SEM}(t_i)} - \sum_{k=1}^{8} \frac{\log_{10}(\lambda_k)^2}{8}, \tag{3.13}$$

where the first summation is the chi-squared function, which quantifies the likelihood of the data given the model, and the second summation relates to the prior probability to select those parameters in the absence of any data. The variable C is a normalization constant that does not depend upon any of the model parameters, and it can be ignored for the purpose of model comparison or likelihood maximization. For small data sets, the second term will dominate, but as more data is collected, the first summation will take on greater importance.

We fit all four possible models (each with a different input-dependent transition rate) by maximizing the likelihood of the model given the data, as specified in Eq. 3.13. For this, we use standard optimization routines, including MATLAB's *fminsearch*, which attempts to find the unconstrained minimum of the objective function using a Nelder-Mead simplex algorithm [109, 110]. We have also utilized a customized simulated annealing algorithm [111]. Knowing the signs of all parameters, we converted them to positive values, and we conducted all searches in logarithmic space. This choice is advantageous because (i) it allows us to improve efficiency by removing need to check and satisfy parameter positivity constraints, and (ii) it allows for parameter step magnitudes to be relative to their current values, even when relative magnitudes of parameters are unknown at the start of the fitting process. To increase our chances of finding a global maximum of the likelihood function (and to avoid getting stuck in local maxima associated with particular parameter guesses), we restarted all searches using thirty different random initial guesses per model. We selected the local maximum that provide the greatest likelihood of the model, given the data. Figure 3.3 shows the result for such a fit using the "correct" model (i.e., Model 2) in which the rate k_{23} was assumed to be time-dependent. The corresponding best-fit parameter set for Model 2 using the ODE fit was:

$$\mathbf{\Lambda_{Fit}} = [1.7254,\ -1.0095,\ 0.1070,\ 0.0998,\ 3.0449,\ 2.2485,\ 1.6607,\ 0.2360]. \tag{3.14}$$

These parameters are very different from the "true" parameters given in Eq. 3.7, suggesting that although the ODE analysis provides excellent fits, it is insufficient for identification of the parameters.

3.4.3 Parameter Uncertainty Estimation for ODE Models

When maximizing the ODE-based likelihood of the model, given the data, we saw that it was possible to find many combinations of models and parameters that match exceptionally well to the simulated data. For example, Models 2 and 4 fit equally well to the data generated from Model 2, meaning that the ODE analysis was insufficient for model discrimination. We next explored

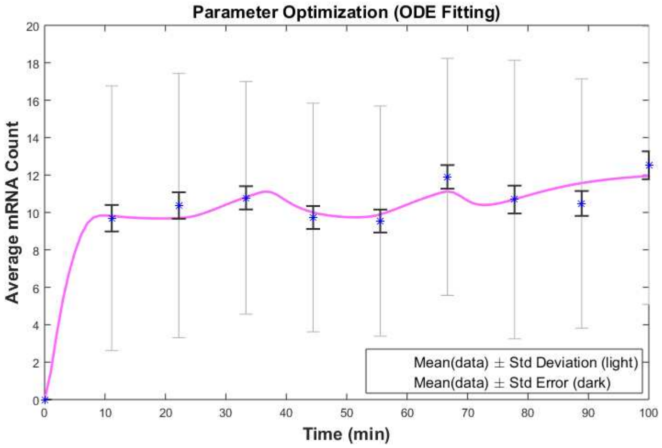


Figure 3.3: Mean gene expression as a function of time for Model 2 after performing optimization to determine the parameter set with the best fit to the averaged data. The light error bars represent the mean of the data ± one standard deviation. The dark error bars represent the mean of the data ± the standard error (100 independent samples per time point). The best ODE-fit parameter set for this model is given in Eq. 3.14.

the uncertainty in parameter space in order to understand how well the averaged experimental data was able to constrain the parameters of the chosen model. To do this, we implemented the Metropolis-Hastings (MH) algorithm (see Section 3.1.1) to explore posterior parameter uncertainty in the model, given the simulated experimental data and the likelihood function in Eq. 3.13.

For our implementation of the MH search, we chose a proposal function in which we select each parameter with probability of 0.5, and, for those selected, we change their logarithmic value by adding a Gaussian-distributed perturbation with a mean of zero and a standard deviation of 0.2. In MATLAB, this can be scripted as:

`x = x + 0.2*randn(size(x)).*(rand(size(x))>0.5)`,

where ‘x’ is the vector of parameters in logarithmic space. For the MH search of parameter space, we used a total chain length of 750,000 parameter sets with a burn-in of 250,000 ignored initial parameter steps and a thin rate of ten ignored for every one recorded parameter set (see Section 3.1.1). Figure 3.4 shows a representative MH trajectory on the plane of parameters k_{r2} and γ for Model 2.

Since the MH algorithm is random, each finite length run of this algorithm produces different results. In practice, the MH algorithm should be run until convergence (e.g., until multiple independent chains with different initial starting parameter guesses converge, within some specified metric and tolerance). Our analysis to compare the ODE model to the simulated data yielded the following average parameter values:

$$\mathbf{\Lambda}_{\mathrm{MH}} = [7.2673,\ 0.1012,\ 0.3457,\ 0.2072,\ 7.2478,\ 0.0706,\ 2.1825,\ 0.0923]$$

and the following large covariances for the parameters in logarithmic space:

$$\Sigma_{\mathbf{\Lambda}} = \begin{bmatrix} 1.5475 & 0.3667 & 0.0827 & -0.0145 & -0.3329 & -0.0032 & 0.5451 & -0.2689 \\ 0.3667 & 1.9240 & -0.0690 & -0.1318 & -0.1888 & -0.2652 & 0.9059 & -0.6151 \\ 0.0827 & -0.0690 & 2.3328 & -0.3474 & 0.3830 & -0.3328 & 0.2021 & -0.3072 \\ -0.0145 & -0.1318 & -0.3474 & 2.7010 & 0.1564 & 0.1150 & -0.6002 & 1.5233 \\ -0.3329 & -0.1888 & 0.3830 & 0.1564 & 0.3285 & 0.0851 & -0.1926 & 0.2883 \\ -0.0032 & -0.2652 & -0.3328 & 0.1150 & 0.0851 & 1.5912 & -0.1730 & 0.6291 \\ 0.5451 & 0.9059 & 0.2021 & -0.6002 & -0.1926 & -0.1730 & 1.6170 & -0.9621 \\ -0.2689 & -0.6151 & -0.3072 & 1.5233 & 0.2883 & 0.6291 & -0.9621 & 1.6560 \end{bmatrix}. \tag{3.15}$$

In Fig. 3.4, the ellipse depicts the 90% confidence interval in the parameters, with the (dubious) assumption that the posterior parameter distribution is a multivariate lognormal distribution with the mean equal to $\mathbf{\Lambda}_{\mathrm{MH}}$ and a covariance matrix equal to $\Sigma_{\mathbf{\Lambda}}$. Clearly, with the data available, *the model is not well identified despite excellent fits to the average data.* This is an illustration of the effect of model sloppiness, where parameters are highly uncertain in some directions, yet potentially well-constrained in other directions within parameter space [112,113]. In Section 3.4.5, we explore how stochastic analyses can dramatically improve model identification, even when using the same model and the same data.

3.4.4 Simulation of Stochastic Gene Expression Models

Next, we simulate many stochastic trajectories for the gene regulation model, thus generating potential single-cell mRNA distributions from the model. To do this, we define the stoichiometry matrix and the propensity functions for each of the four possible models. Unlike the standard

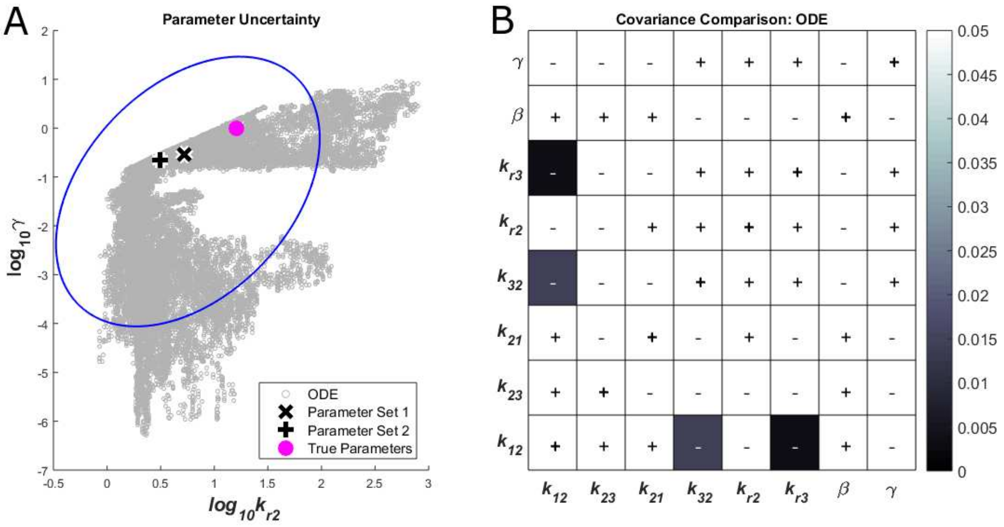


Figure 3.4: (A) Parameter uncertainty in logspace for parameters k_{r2} and γ, as determined by applying the Metropolis-Hastings algorithm to the ODE likelihood function (Eq. 3.13). The ellipse indicates a 90% confidence interval, assuming a log Gaussian posterior. The magenta circle indicates the true k_{r2} and γ parameter values. Two similarly good parameter sets were selected from this chain: $log_{10}(\mathbf{\Lambda_1}) = [0.0705,\ -2.2859,\ -0.1497,\ -0.3444,\ 0.7181,\ -1.9591,\ -1.1805,\ -0.5304]$ and $log_{10}(\mathbf{\Lambda_2}) = [0.1797,\ 0.5551,\ -0.7058,\ -1.1098,\ 0.4949,\ 0.3303,\ 0.6639,\ -0.6531]$. For these parameters sets, k_{r2} and γ are represented by the black **x** and **+**, respectively. (B) Comparison of the *thresholded* covariances for each parameter combination from the ODE MH search, with the covariance values thresholded at ± 0.05. All covariance values whose magnitudes exceed 0.05 are shown in white. The **(+)** and **(-)** symbols indicate positive and negative covariances, respectively.

SSA, one of the propensity functions in these models is a time-varying function. In this case, it is necessary to adapt the SSA in a manner similar to that discussed in [114, 115]. This modification corresponds to adding a fast reaction to the system, which has a large propensity function ($w_1 = 100/\mathrm{min}$), but which does not change the state of the system ($\mathbf{S}(:,1) = \mathbf{0}$). We ran 1000 SSA trajectories for each of the four possible models, and we made histograms of the number of mRNA in different cells at each point in time (using the resulting best fit parameters from Section 3.4.2).

A single SSA trajectory is shown Fig. 3.5A. In Fig. 3.5B, we plot the average of multiple sets of 1000 SSA simulations versus time along with the ODE solution. In addition, Fig. 3.5C shows different FSP distributions for the two parameter sets (given in the caption for Fig. 3.4) and the simulated data at the same, specific time point. Despite producing excellent fits to the averaged data (Fig. 3.5B), neither parameter set captures the bimodal behavior of the data, as shown in Fig. 3.5C. This again shows that fitting the set of ODEs was not a sufficient strategy to extract all

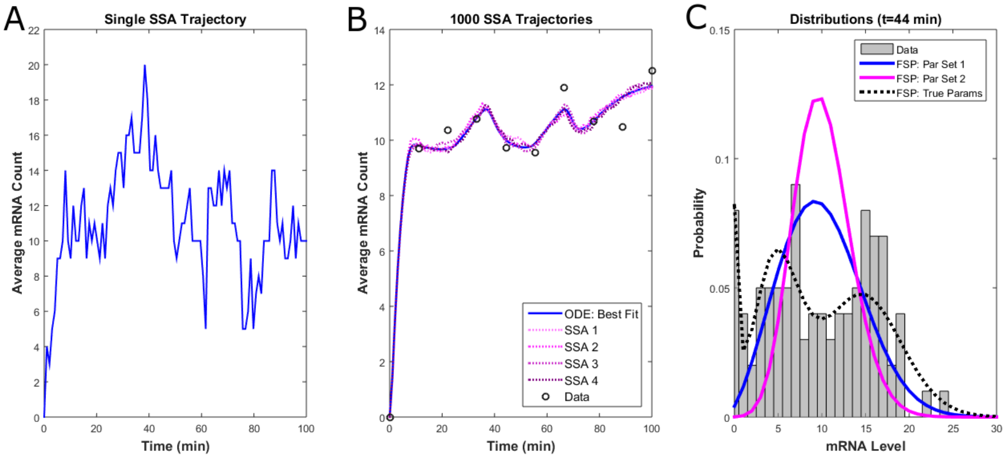


Figure 3.5: (A) Results from a single SSA trajectory for Model 2 using the best fit parameters, $\mathbf{\Lambda_{Fit}}$, obtained in Section 3.4.2. (B) Mean gene expression as a function of time from multiple runs of 1000 SSA trajectories using the best fit parameters compared to the deterministic solution and the experimental data. (C) Full distributions at 44 minutes for the two parameter sets, $\mathbf{\Lambda_1}$ and $\mathbf{\Lambda_2}$, compared to the data and the exact distribution for the true parameters at the same time point. The true parameters are $\mathbf{\Lambda_{True}} = [0.20,\ -0.076,\ 0.050,\ 0.050,\ 16.0,\ 5.0,\ 0.10,\ 1.0]$ or equivalently $\mathbf{log_{10}}(\mathbf{\Lambda_{True}}) = [-0.6990,\ -1.1192,\ -1.3010,\ -1.3010,\ 1.2041,\ 0.6990,\ -1.0000,\ 0]$.

available information from our simulated single-cell data set, and the result suggests that fitting full distributions could lead to improvements in parameter identification [22]. We examine this possibility further in the next subsection.

3.4.5 Parameter Uncertainty Estimation for FSP Models

To improve upon the parameter uncertainty estimation results, we next utilized the Finite State Projection (FSP) approach to re-explore the uncertainty in parameter space, given the single-cell data. To accomplish this, we write the CME in the matrix ODE form:

$$\frac{d\mathbf{P}}{dt} = \mathbf{A}(t)\mathbf{P}(t), \tag{3.16}$$

where $\mathbf{P} = [p_1, p_2, \ldots]$ is the vector of probabilities of each state of the stochastic process, and $\mathbf{A}(t)$ is known as the *infinitesimal generator matrix*. See [55, 57, 58, 97] for examples on the construction and evaluation of this FSP formulation (Eq. 3.16) and the GITHUB folder at "https://github.com/MunskyGroup/WeberPB/Matlab_Codes" for specific FSP codes for this section. Equation 3.16 can then be integrated in time to compute full probability distributions for the gene regulation dynamics. We verified that the FSP analysis produced distributions that matched to the SSA results in the previous section (Fig. 3.5C, compare bars and dashed line). With the FSP-computed probability distributions, the likelihood of the data, given the model, is easily computed using the expression:

$$\log\left(L_{\mathrm{FSP}}(\mathbf{D}|\mathbf{\Lambda})\right) = \sum_{i=1}^{N_t}\sum_{j=0}^{N_r} n_{ij}\log(p(j|t_i,\mathbf{\Lambda})), \tag{3.17}$$

where N_r is the maximum number of mRNA observed in a single cell, n_{ij} is the number of cells with exactly j mRNA at the i^{th} time point, and $p(j|t_i,\mathbf{\Lambda})$ is the probability to have j mRNA at that time, given the model parameters in $\mathbf{\Lambda}$, as discussed in [55, 58]. The corresponding log-likelihood of the model, given the data, can be written (using our previous definition of the prior) as:

$$\log\left(L_{\mathrm{FSP}}(\mathbf{\Lambda}|\mathbf{D})\right) = C - \log(P(\mathbf{D})) + \sum_{i=1}^{N_t}\sum_{j=0}^{N_r} n_{ij}\log(p(j|t_i,\mathbf{\Lambda})) - \sum_{k=1}^{8}\frac{\log_{10}(\lambda_k)^2}{8}, \tag{3.18}$$

where C is a constant that is independent of the parameters. With this new likelihood function, we again implemented the MH algorithm to explore parameter uncertainty in the model, given the full information within the simulated single-cell data. In comparison to the parameter uncertainty estimation using the ODE analysis, our analysis using the FSP approach yielded substantially lower covariances for the parameters, as seen in the logarithmic covariance matrix:

$$\Sigma_{\mathbf{\Lambda}} = \begin{bmatrix} 0.0079 & 0.0049 & 0.0084 & -0.0008 & 0.0099 & 0.0105 & 0.0007 & 0.0100 \\ 0.0049 & 0.5976 & 0.0097 & -0.0133 & 0.0010 & 0.0022 & 0.0837 & 0.0008 \\ 0.0084 & 0.0097 & 0.0148 & 0.0021 & 0.0036 & 0.0042 & 0.0026 & 0.0034 \\ -0.0008 & -0.0133 & 0.0021 & 0.0089 & -0.0064 & -0.0067 & 0.0022 & -0.0067 \\ 0.0099 & 0.0010 & 0.0036 & -0.0064 & 0.0314 & 0.0328 & -0.0012 & 0.0322 \\ 0.0105 & 0.0022 & 0.0042 & -0.0067 & 0.0328 & 0.0348 & -0.0011 & 0.0336 \\ 0.0007 & 0.0837 & 0.0026 & 0.0022 & -0.0012 & -0.0011 & 0.0223 & -0.0014 \\ 0.0100 & 0.0008 & 0.0034 & -0.0067 & 0.0322 & 0.0336 & -0.0014 & 0.0330 \end{bmatrix}. \quad (3.19)$$

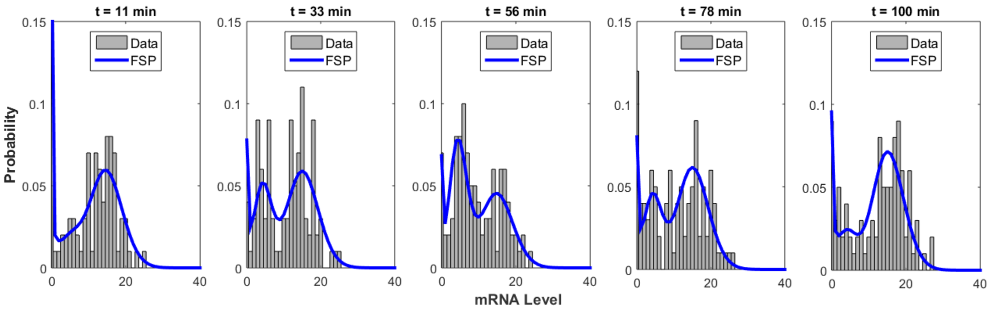
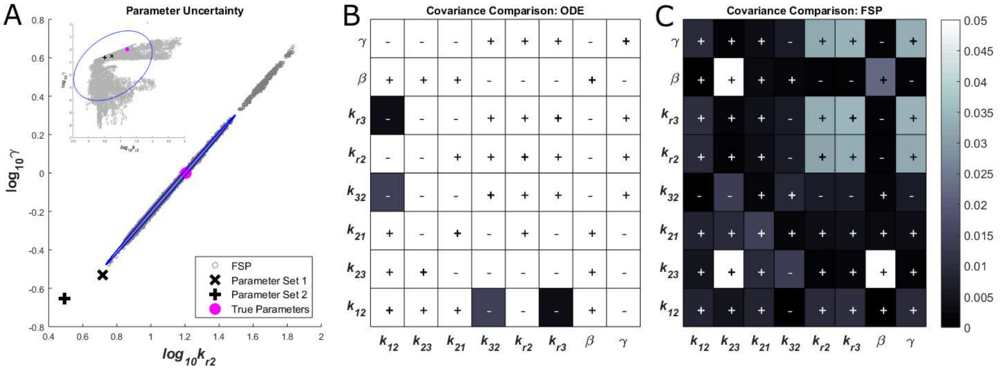


Figure 3.6: Data distributions compared with FSP distributions for Model 2 (using the FSP best fit parameter values) at specific time points.

The FSP distribution results are presented in Figs. 3.6 and 3.7. Figure 3.6 shows that the FSP distributions resulting from the identified parameters match closely to the data distributions at the time points presented and capture the bimodal behavior of the data. Unlike the ODE analysis, the FSP was able to recover the correct model (i.e., Model 2) from its fit to the simulated data. Moreover, Fig. 3.7 shows that there is substantially less parameter uncertainty when the Metropolis-Hastings search is performed with the FSP instead of the ODE analysis. Figure 3.8 shows the true parameter values compared to the average estimated parameter values for Model 2 from the ODE and FSP searches of the parameter space. For nearly all of the parameters, the FSP search yields substantially closer estimates of the true parameter values and has much lower



B: Covariance Comparison: ODE

γ	-	-	-	+	+	+	-	+
β	+	+	+	-	-	-	+	-
k_{r3}	-	-	-	+	+	+	-	+
k_{r2}	-	-	+	+	+	+	-	+
k_{32}	-	-	-	+	+	+	-	+
k_{21}	+	-	+	-	+	-	+	-
k_{23}	+	+	-	-	-	-	+	-
k_{12}	+	+	+	-	-	-	+	-
	k_{12}	k_{23}	k_{21}	k_{32}	k_{r2}	k_{r3}	β	γ

C: Covariance Comparison: FSP

γ	+	+	+	-	+	+	-	+
β	+	+	+	+	-	-	+	-
k_{r3}	+	+	+	-	+	+	-	+
k_{r2}	+	+	+	-	+	+	-	+
k_{32}	-	-	+	+	-	-	+	-
k_{21}	+	+	+	+	+	+	+	+
k_{23}	+	+	+	-	+	+	+	+
k_{12}	+	+	+	-	+	+	+	+
	k_{12}	k_{23}	k_{21}	k_{32}	k_{r2}	k_{r3}	β	γ

Figure 3.7: (A) Parameter uncertainty for k_{r2} and γ, as determined using the FSP likelihood function (Eq. 3.18) in the Metropolis-Hastings analysis of Model 2. The ellipse indicates a 90% confidence interval. The magenta circle indicates the true k_{r2} and γ values. (A Inset) To provide a scale reference, the FSP-MH results are compared to the ODE search results for the same data set. (B,C) Comparison of the *thresholded* covariances for each parameter combination from the ODE MH search (B) and the FSP MH search (C), using the thresholding procedure described in the caption of Fig. 3.4 caption. The **(+)** indicates a positive covariance and **(-)** indicates a negative covariance.

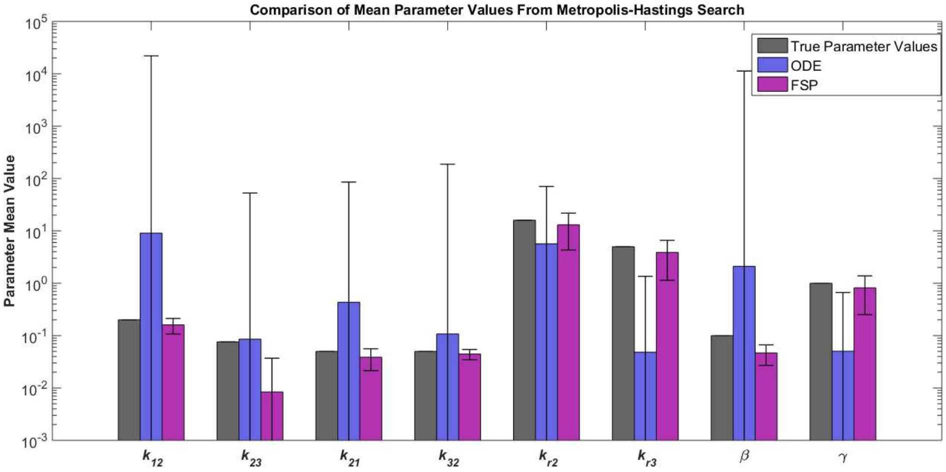


Figure 3.8: Bar plot showing the true parameter values compared to the mean parameter values from the ODE MH search and the FSP MH search for Model 2. The error bars represent the mean parameter value ± one standard deviation.

standard deviations for the parameter estimates as compared to ODE-based search. The only exception to this observation, for parameter $k_{23}(0)$, is easily explained by the fact that this reaction (Eq. 3.2) is saturated and, therefore, unobservable for much of the system trajectory. It is also worth noting that, in addition to providing better parameter estimates and lower uncertainty, the FSP-based analysis required far shorter MH chains to explore parameter space before converging around the global minimum.

3.4.6 Directions of Parameter Uncertainty

To quantify and compare the model sloppiness after parameter estimation using the ODE and FSP analyses, we examined the eigenvectors and eigenvalues of the MCMC results (in logarithmic space). The square-root of the largest eigenvalue and its corresponding eigenvector quantify the magnitude and direction of the greatest component of parameter uncertainty (also known as the first principal component or PC). The second eigenvalue/eigenvector pair characterizes the next most variable direction, and so on. Figure 3.9A shows the magnitudes of the eight ranked PCs for the ODE (red) and FSP (black) analyses. The solid lines in Fig. 3.9A correspond to identification using the initial condition (S_1 and $R = 0$) as presented above, and the dashed lines correspond to using a different initial condition defined by the steady-state distribution in the absence of external input. For this particular combination of input signal and sampling frequency, the two different initial conditions yield similar results for the contrast between the ODE and FSP analyses. However, we note that both the ODE and the FSP analyses are slightly more variable when the initial condition is defined by the steady-state behavior in the absence of the activating signal.

Figures 3.9B-E show the directions of all eight PCs on the $\log_{10}(k_{r2}) \times \log_{10}(\gamma)$ plane for the FSP and ODE analyses with the two different initial conditions. For the FSP analysis with the zero-activation initial condition (Fig. 3.9 B), the first PC explains 86$\%$ of the total uncertainty, and only three PCs are needed to explain more than 95$\%$ of the total uncertainty. Moreover, parameters k_{r2} and γ are linearly related to one another (i.e., their logarithms co-vary with a ratio of one) in each of the first three PCs, and Fig. 3.9B shows almost no uncertainty except on a single line with unit slope (Fig. 3.9C shows the same trend for the other initial condition). In contrast, six of the eight of the possible PCs are required in order to explain the total uncertainty in the ODE analysis, and Figs. 3.9C,E show that all of these PCs have significant variation in the $\log_{10}(k_{r2}) \times \log_{10}(\gamma)$ plane, for either initial condition.

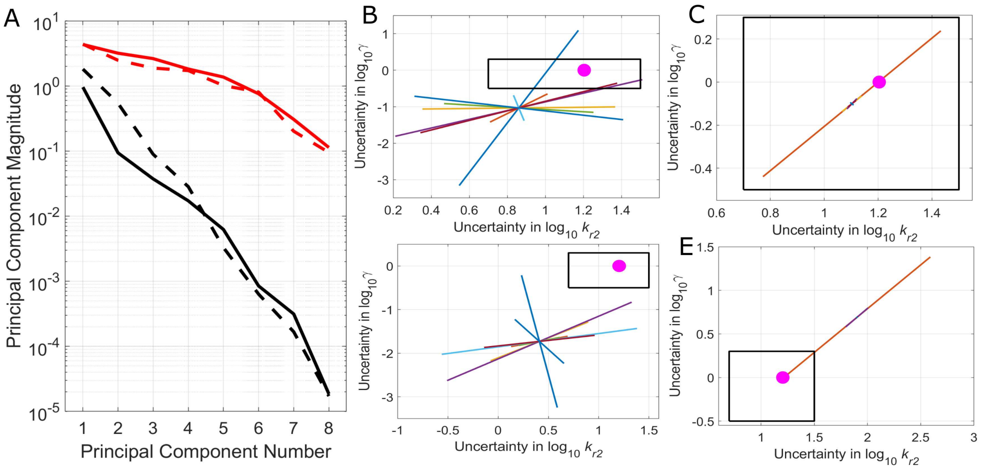


Figure 3.9: Principal component analysis of the posterior uncertainty after parameter estimation. (A) The magnitude of uncertainty (i.e., parameter variances) for the FSP (black) and ODE (red) analyses, ranked from the least to the best constrained directions in parameter space. Solid lines denote results using the initial conditions described in the text, and dashed lines denote the identification results assuming that the initial conditions are determined by the equilibrium condition in the absence of external signal. (B,C) Directions and magnitudes (± two standard deviations) of the principal components on the $\log_{10}(k_{r2})$ versus $\log_{10}(\gamma)$ plane using the (B) FSP analysis or (C) ODE analysis. Panels (B,C) use the initial conditions described in the text, and (D,E) use the zero-input steady state as the initial condition. For scale comparison of scales, the black box is the same size in every panel, and the magenta circle denotes the true parameter set.

3.4.7 Importance of Experiment Design

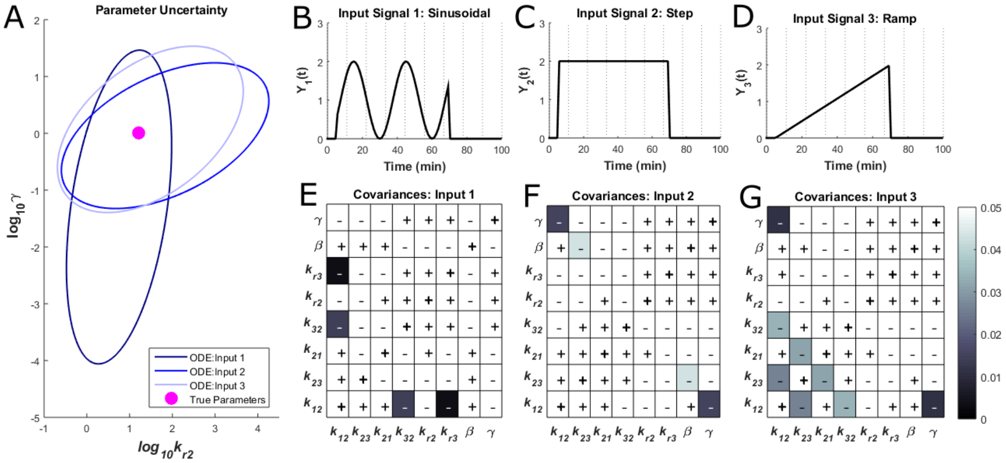
Up to this point, we have demonstrated that the type of analysis chosen (i.e., an ODE- or FSP-based likelihood function) to examine experimental data is crucial in model and parameter identification. We have also observed that the type of initial condition can have a smaller, but still significant quantitative effect on parameter identification. We now briefly explore how different experimental designs can affect parameter identification results. To demonstrate this concern, we fit the data and performed MH searches of parameter space for multiple different experiments (i.e., different input signals). Specifically, we compared the original input signal, given in Eq. 3.1, with two additional input signals, which are as follows:

$$Y_2(t) = \left\{ \begin{array}{ll} 0 & \text{for } (t \leq 5)(t \geq 70) \\ 2 & \text{for } t \in (5, 70) \end{array} \right\}, \tag{3.20}$$

$$Y_3(t) = \left\{ \begin{array}{ll} 0 & \text{for } (t \leq 5)(t \geq 70) \\ \left(\frac{2}{65}t\right) - \left(\frac{2}{13}\right) & \text{for } t \in (5, 70) \end{array} \right\}, \tag{3.21}$$

where both additional input signals begin at time $t = 5$ min and end at time $t = 70$ min, similar to $Y_1(t)$.

Figures 3.10 and 3.11 illustrate how the three different experimental designs (i.e., three different input signals) lead to dramatically different parameter uncertainties, using the ODE and the FSP likelihood functions, respectively. In all figures, we assume the exact same amount of data, but the effects on which parameters are tightly identified and which remain uncertain is non-trivially dependent on the input signal. Moreover, input signals that are best for one type of analysis may not be the best choice for a different analysis (for example, compare Figs. 3.10G to 3.11G).



Covariances: Input 1

	k_{12}	k_{23}	k_{21}	k_{32}	k_{r2}	k_{r3}	β	γ
γ	-	-	-	+	+	+	-	+
β	+	+	+	-	-	-	+	-
k_{r3}	-	-	-	+	+	+	-	+
k_{r2}	-	-	+	+	+	+	-	+
k_{32}	-	-	-	+	+	+	-	+
k_{21}	+	-	+	-	+	-	+	-
k_{23}	+	+	-	-	-	-	+	-
k_{12}	+	+	+	-	-	-	+	-

Covariances: Input 2

	k_{12}	k_{23}	k_{21}	k_{32}	k_{r2}	k_{r3}	β	γ
γ	-	-	-	-	+	+	+	+
β	+	-	-	-	+	+	+	+
k_{r3}	-	-	-	-	+	+	+	+
k_{r2}	-	-	+	-	+	+	+	+
k_{32}	-	+	+	+	-	-	-	-
k_{21}	+	+	+	+	+	-	-	-
k_{23}	+	+	+	+	-	-	-	-
k_{12}	+	+	+	-	-	-	+	-

Covariances: Input 3

	k_{12}	k_{23}	k_{21}	k_{32}	k_{r2}	k_{r3}	β	γ
γ	-	-	-	-	+	+	+	+
β	+	+	-	-	+	+	+	+
k_{r3}	+	-	-	-	+	+	+	+
k_{r2}	-	-	+	-	+	+	+	+
k_{32}	-	+	+	+	-	-	-	-
k_{21}	+	-	+	+	+	-	-	-
k_{23}	-	+	-	+	-	-	+	-
k_{12}	+	-	+	-	-	+	+	-

Figure 3.10: (A) Uncertainty in k_{r2} and γ given three different input signals, as determined by the ODE-based MH analysis of Model 2. The ellipses indicate a 90% confidence interval. The magenta circle indicates the true k_{r2} and γ parameter values. (B,C,D) Three different input signals and (E,F,G) the corresponding comparison of the *thresholded* covariances for each parameter combination from the ODE MH analysis. The **(+)** indicates positive covariances and **(-)** indicates negative covariances.

To further demonstrate how important experimental design considerations are to obtain accurate prediction of gene regulation dynamics, we make model predictions of FSP distributions at different times and for multiple different experiments (i.e., different input signals). In these predictions, we use the best-fit parameter sets from the FSP-based searches for all four models (each

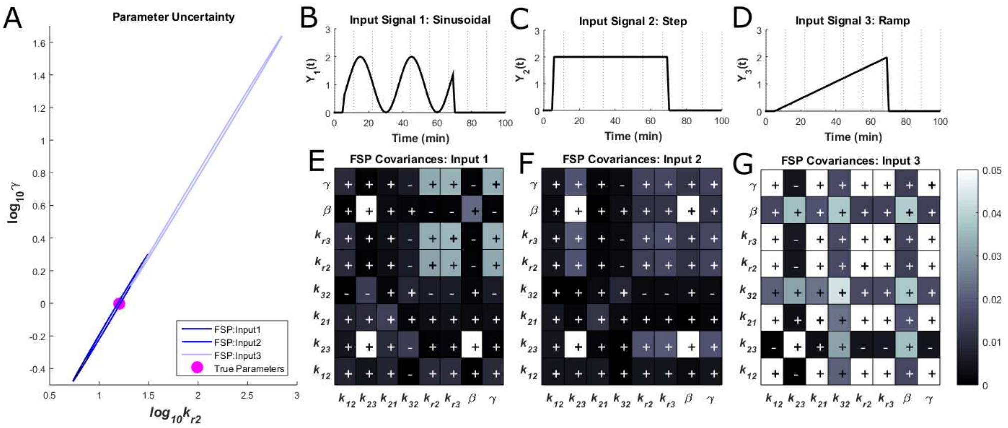


Figure 3.11: (A) Uncertainty in k_{r2} and γ given three different input signals, as determined by the FSP-based MH analysis of Model 2. The ellipses indicate a 90% confidence interval. The magenta circle indicates the true k_{r2} and γ parameter values. (B,C,D) Three different input signals and (E,F,G) the corresponding comparison of the *thresholded* covariances for each parameter combination from the FSP MH analysis. The **(+)** indicates positive covariances and **(-)** indicates negative covariances.

with a different time-dependent transition rate and obtained using the original input signal given in Eq. 3.1). The potential input signals are $Y_2(t)$ (Eq. 3.20), $Y_3(t)$ (Eq. 3.21), and $Y_4(t)$, given by:

$$Y_4(t) = \left\{ \begin{array}{ll} 0 & \text{for } (t \leq 5)(t \geq 70) \\ 1 - \cos\left(\frac{2\pi t}{15}\right) & \text{for } t \in (5, 70) \end{array} \right\}. \tag{3.22}$$

Figure 3.12 shows predictions for these 3 potential input signals (left, middle, right columns) for two potential time points (middle, bottom rows), assuming the 4 different models that were identified using the original experiment (different colors). By examining these predictions, we can gain insight into which experiments might be more informative. For example, at t = 30 min (middle row), an experiment with input signal $Y_2(t)$ (left column) or $Y_3(t)$ (middle column) would enable us to differentiate easily between the models, whereas input signal $Y_4(t)$ (right column) is unlikely to differentiate between the models. Furthermore, measuring the output at t = 30 min is clearly more informative than at t = 100 min (compare Figs. 3.12D-F with 3.12G-I).

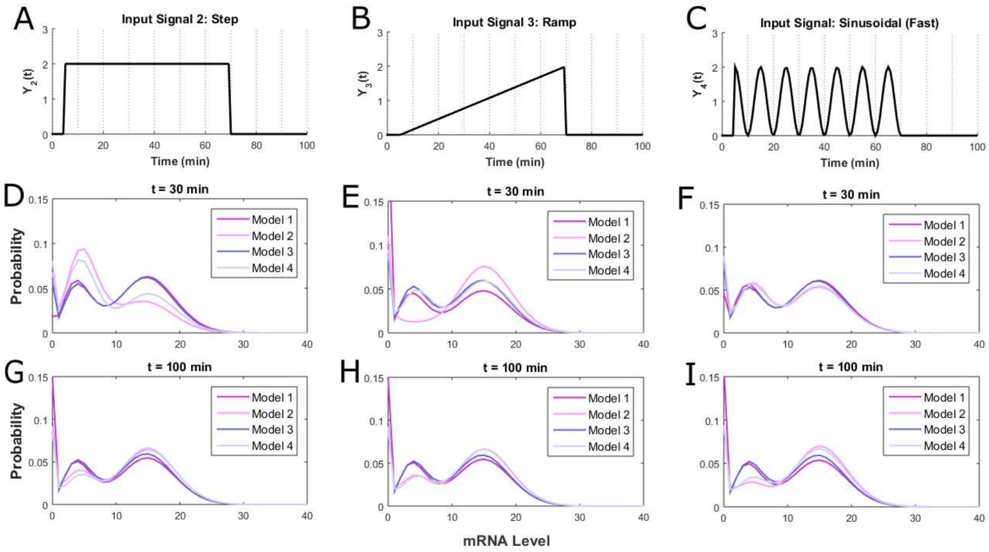


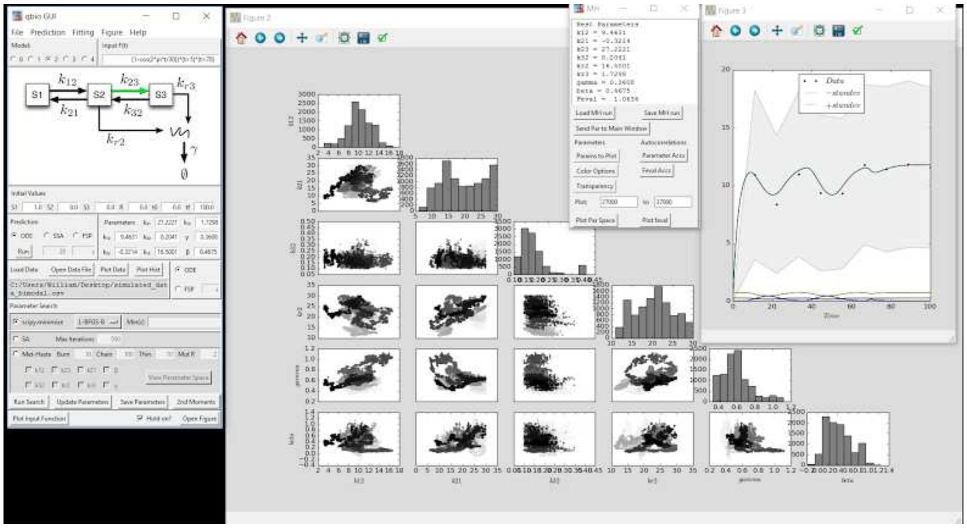
Figure 3.12: Predictions of the FSP distributions for all four model numbers at two different time points for three different experiments (with input signals shown in Eq. 3.20 - 3.22). Each model uses the best parameter set found during the fit to the original input signal in Eq. 3.1.

3.5 Summary and Conclusions

In this article, we explored the application of various computational techniques (such as ODE analyses [32–37], stochastic simulations [38–49], and finite state projections [51–60]) to understand the dynamics of single-cell gene regulation. In particular, we focused on a class of simplified multi-state gene regulation models that has previously been fit to experimental single-cell data [1,2,22], and we used a combination of simulated single-cell data and different modeling procedures to explore how well one could identify models and parameter uncertainties from discrete stochastic data. Additionally, we discussed the importance of experiment design and illustrated how different initial conditions or experiments can be more informative than others for model identification. In addition to changing the initial conditions or input conditions, additional information can be gained from system response by adjusting the times at which measurements are taken. Although there have been many recent developments toward model-driven experiment design for ODE analyses [99, 116], the development of experiment design methods for stochastic gene regulatory systems remains an active field of investigation [99, 117]. Moreover, our analyses show that fitting ODEs to dynamical data may not be adequate for parameter identification and discerning parameter uncertainty (even when models are relatively simple), and we discussed the fact that fitting full distributions (e.g., using an FSP likelihood function) may lead to substantial improvements, even when considering the same model and same data. These results argue that the integration of discrete stochastic models and single-cell data could open new possibilities for improved model identification, even in cases where traditional ODE analyses have failed due to excessive parameter sloppiness.

In our current study, we focused on the analysis of mRNA levels as can be measured precisely with single-molecule resolution and fast temporal resolution using single-molecule Fluorescence *in situ* Hybridization [1, 10, 11, 22]. Similar tools and stochastic model can also be applied to the analysis of protein variations, although such analyses must contend with additional complications, such as slow protein translation or activation dynamics that can filter out the fast transcriptional responses [94] or added background noise (e.g., cellular autofluorescence) that must be convolved with the model-generated protein signal [118]. Despite these issues, single-cell flow cytometry data have been successfully combined with stochastic models to generate accurate predictive models for single-cell protein distributions [118, 119]. We assumed the possibility of rich and well measured time-varying input signals, which can be temporally controlled through alteration of environmental conditions and quantified with time-lapse fluorescence microscopy [1, 2, 105–107].

To make these approaches more accessible, we have developed a GUI, which is freely available on GITHUB in the folder “https://github.com/MunskyGroup/WeberPB/GUI”. The GUI, shown in Fig. 3.13, provides an opportunity for users to specify any input and model number and to perform the analyses outlined throughout, evaluate results, and generate figures. In addition, the GUI enables the automatic generation of simulated data using SSA for user-specified parameter sets and models, or to import a previously generated data file to perform the analyses outlined in this exercise. The GUI requires PYTHON and the packages: NUMPY, SCIPY, and MATPLOTLIB.



```
C:/Users/William/Desktop/simulated_dat
a_bimodal.csv
```

```
Best Parameters
k12 = 9.4431
k21 = -0.3214
k23 = 27.2221
k32 = 0.2041
kr2 = 16.5001
kr3 = 1.7258
gamma = 0.3608
beta = 0.4675
Feval = 1.0836
```

Figure 3.13: Screenshot of the provided GUI, which can be utilized to perform all analyses outlined throughout, including generating simulated data sets and producing desired plots.

Chapter 4

Deconstructing the kinetic rates of binding and site-unwinding in structure-mediated nucleic acid interactions[3]

4.1 Introduction

Proper cellular function relies on proteins, RNA, and DNA adopting the correct structural conformations to support biological activity. Dynamic changes in the structures of nucleic acid molecules allow for a wide variety of interactions that are critical to cellular processes. One particular example is DNA unwinding, or the process of the melting of a DNA strand that makes it stably single-stranded, which exposes a target nucleotide sequence to other molecules, such as enzymes, proteins, and various additional molecules involved in DNA replication, repair, and transcription [27–29]. These DNA structural changes and their associated interactions are sensitive to changes in multiple variables, such as DNA supercoiling, ionic strength, and temperature [120, 121]. Cellular processes can also dynamically affect the local, or even global, DNA structure they act upon, promoting or inhibiting further interactions [121–124]. Despite the importance of DNA structural kinetics and interactions to many cellular processes, there is insufficient information on their dynamics.

Controlling interactions between biomolecules and target DNA sites is also of increasing importance within the biotechnology community. New gene editing techniques, such as CRISPR-

---

[3] The work discussed in this chapter has been done in collaboration with Shane Scott, Cynthia Shaheen, Brian Munsky, and Sabrina R. Leslie. Experimental data modeled in this work has been provided by Shane Scott and Sabrina R. Leslie, and are published in Scott, *et al.* (2018) [3] and Scott, *et al.* (2019) [4]. Computational analyses and modeling presented here were completed by me. In this manuscript, the text has been authored by both Shane Scott and me. I have primarily authored the introduction and text regarding the modeling and results, while Shane Scott has primarily authored the text related directly to physical interpretations. While portions of the models developed and used here have been published in Shaheen, *et al.* (2022) [26], much of the work presented here is currently unpublished.

Cas9, employ guide-RNA oligonucleotides (oligos) that bind to complementary DNA sequences as a first step in genetic modification processes within living cells [86, 87]. Beyond gene editing applications, understanding DNA structural transitions, such as unwinding, and their subsequent interactions could have a profound impact on the development of oligonucleotide therapeutics [12, 125]. As such, these techniques would benefit from a more sophisticated understanding of DNA unwinding and rewinding dynamics as functions of supercoiling, temperature, or ionic conditions, and their role in mediating nucleotide interactions.

Previous studies have performed experimental and computational analyses to explore conformational changes of RNA or DNA molecules in relation to environmental conditions or mechanical stress, many of which utilize molecular dynamics (MD) simulations [126–141]. Liangruksa and Wongwises (2018), for example, applied Kirkoff's rod model to evaluate the elastic behavior and DNA conformational changes resulting from an increase in temperature [130]. Kruithof and van Noort [138] and Bian, *et al.* [139] used Markov Models to explore the conformational dynamics of DNA, where the transitionary events were assumed to be stochastic, dependent only upon the previous state of the system [141]. In addition, Ghosh, *et al.* [134] utilized Monte Carlo (MC) and MD simulations to analyze the effects of ionic strength on DNA conformational changes. Works that analyze the binding dynamics of DNA have also been performed [142–148]. For example, Palermo, *et al.* [149] explored conformational changes and binding dynamics in relation to the CRISPR-Cas9 gene editing system using MD simulations. Additionally, Zarrabi, *et al.* [150] used hidden Markov models to investigate conformational changes and binding dynamics related to the formation of the TBP-NC2-DNA complex. Studies have also been performed on analyzing and modeling DNA microarray data, including using Markov Chain Monte Carlo (MCMC) simulations for parameter estimation [151–153]. While these previous studies have simulated or modeled the effects of temperature, supercoiling, or ionic strength on DNA structural transitions or subsequent binding dynamics independently, these factors all act upon nucleic acids simultaneously *in vivo*. It is thus necessary to develop a more general model that incorporates all of these variables at the same time to determine their collective importance to DNA structural changes and concurrent interactions.

In this work, we present a simple computational model that quantitatively captures both the structural transition of double-strand DNA (dsDNA) to an unwound (stable single-stranded) structure and the binding dynamics of a complementary DNA-oligonucleotide probe. We investigate these dynamics as a function of several variables: temperature, salt concentration, and DNA supercoiling. The model consists of a three-state Markov Chain that describes the unwinding and rewinding dynamics of the DNA double-helix and subsequent oligonucleotide probe binding dynamics. We fit our model to experimental data of DNA unwinding and probe binding to a particular site on plasmid DNA [3, 4] using a modified genetic algorithm to find parameters that maximize the likelihood of the data, and we use Markov Chain Monte Carlo (MCMC) simulations (i.e., the Metropolis Hastings algorithm [61]) to quantify parameter uncertainties.

4.1.1 Experimental Methods

Recently, Scott, *et al.* (2018, 2019) presented a novel means of observing DNA unwinding and interactions between a known unwinding region and a fluorescently-labeled DNA oligo probe [3, 4]. Using Convex Lens-induced Confinement (CLiC) microscopy, a small portion of a solution containing pUC19 plasmid molecules and oligos complementary to a known pUC19 unwinding region (Site 1, associated with the origin of replication in pUC19) were trapped in 3 μm-diameter, 500 nm-deep wells. In these wells, it was possible to monitor the interactions occurring in the large sample volume by recording multiple short videos, each of a fresh sampling of molecules from the larger reaction volume, thus characterizing system behavior over a long period of time. By counting the number of interacting molecules within a sample as a function of temperature, ionic strength, and supercoiling, it was observed that binding increased with increasing temperature and supercoiling, or with decreasing solution ionic strength [3, 4].

While these works validated the experimental technique, the model used to extract the reaction rate between plasmids and oligo probes and to quantify DNA opening was unable to accurately measure the unwinding and rewinding rates. Previously, it was assumed that plasmid unwinding and rewinding were very slow compared to the timescale of plasmid-probe binding, and these rates were ignored in the mathematical model. Though experimental controls supported these assumptions, the presented binding curves never quite reached stable equilibrium, showing binding to increase slightly with time at late experimental timescales. Such a weak time dependence suggests that unwinding and rewinding may indeed affect the probe-plasmid binding interactions, even if only slightly. Thus, by applying a more complex modeling approach, we hypothesize that it may be possible to simultaneously extract multiple kinetic rates from this data: those of binding, unbinding, unwinding, and rewinding. These dynamic processes occur frequently within living cells, but their values and how they interact with one another are relatively unknown. Quantifying these rates could therefore provide important insight into cellular processes, and increase our understanding of DNA structure and interactions.

4.1.2 Modeling Approach

This work builds upon Scott, *et al.* (2018, 2019) by applying new computational models to investigate the structural dynamics of unwinding and rewinding in Site 1, and the subsequent interaction kinetics between this site and complementary oligo probes. Figure 4.1A illustrates this system using three distinct, reversible conformations (‘closed’, ‘open’, and ‘bound’) of the oligo-plasmid system. In our computational analyses, C corresponds to the closed DNA plasmid with no bound oligo probe (wound structure); O corresponds to the open DNA plasmid and unbound oligo probe (unwound structure); B corresponds to the open, bound oligo-plasmid system (bound structure), and U corresponds to the unbound oligo probe. Here, we examine the possible transition rates between C, O, and B, and determine their dependence upon temperature (T), ionic strength ($[S]$), and/or DNA supercoiling (σ).

When modeling the interacting oligo-plasmid system, we assume a set of four transition events between the closed, open, and bound conformations: plasmid opening, $i = o$, plasmid closing, $i = c$, oligo binding, $i = b$, and oligo unbinding, $i = u$, each with a corresponding transition rate

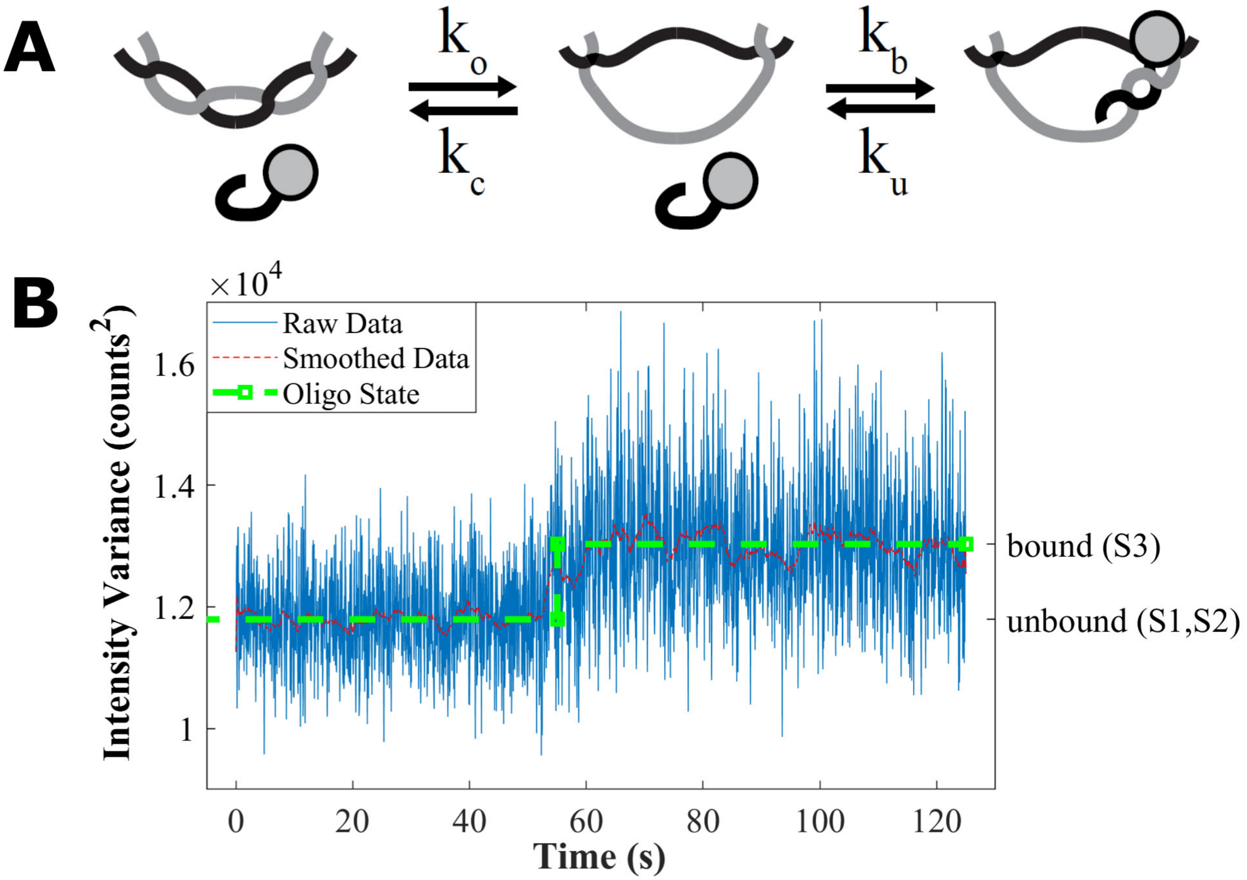


Figure 4.1: A) A schematic of the 3-state Markov Chain used to describe the three conformational structures of the oligo-plasmid system over time. Each transition rate, k_i, is potentially dependent upon temperature, salt concentration, and/or DNA supercoiling. B) Spatial variance vs. time for an experimental oligo-plasmid binding event, including raw and smoothed data, as well as the corresponding conformational structures. The data was obtained at $T = 31°\text{C}$, 22.5 mM total ionic concentration, and a plasmid with $\langle\sigma\rangle = -0.151$.

of k_i. These transitions are assumed to follow the principles of a Markov chain: they are stochastic, with transition rates dependent only upon the current state (conformation) of the system. Here, the plasmids and oligos are assumed to be well-mixed prior to imaging. All transition rates are non-negative, and each could depend upon one or more of the experimental conditions (T, $[S]$, σ). To ensure that the reaction rates are positive in all environments, we define each reaction rate using the form:

$$\begin{aligned} k_i &= \mathrm{e}^{f(T-T_0, S-S_0, \sigma-\sigma_0)}, \\ &= \mathrm{e}^{f(\Delta T, \Delta S, \Delta\sigma)}, \end{aligned} \tag{4.1}$$

where f is an unknown function that we wish to identify from experimental data. For convenience, Eq. 4.1 is defined relative to a reference temperature, $T_0 = 310$ K, reference ionic strength, $[S]_0 = 22.5$ mM, and reference supercoiling, $\langle\sigma_0\rangle = -0.108$. We reasoned that the simplest form for f should be affine linear in each of the independent variables, such that:

$$\begin{aligned} k_i &= \mathrm{e}^{k_{i,0}+\alpha_{i,T}\Delta T+\alpha_{i,S}\Delta S+\alpha_{i,\sigma}\Delta\sigma}, \\ &= \mathrm{e}^{k_{i,0}}\mathrm{e}^{\alpha_{i,T}\Delta T}\mathrm{e}^{\alpha_{i,S}\Delta S}\mathrm{e}^{\alpha_{i,\sigma}\Delta\sigma}. \end{aligned} \tag{4.2}$$

Under this definition, $k_{i,0}$ is the baseline value for each transition rate, which ensures that all transition rates are positive; this baseline value also ensures that the transition rate is non-zero if it is independent of all experimental conditions. The $\alpha_{i,y}$ denotes how strongly the structural transition depends upon the corresponding independent variable (represented by y), with positive or negative values corresponding to either an increase or decrease in that rate (depending upon the sign of the experimental variable), respectively. Null values correspond to an independence of that reaction with respect to the experimental variable. As a special case, if a transition rate is not dependent upon any of the experimental conditions, or if the system is operating at the baseline conditions $(T_0, [S]_0, \sigma_0)$, then the transition rate is reduced to the baseline value, $k_i = \mathrm{e}^{k_{i,0}}$.

We also note that this choice of general definition for the dependence of the rates on the parameters is similar to the Arrhenius equation (at least for the case of small temperature variations), in which the dependence of k_i on temperature would be assumed to be:

$$k_i(T) = A\mathrm{e}^{\left(\frac{-E_a}{RT}\right)}. \tag{4.3}$$

If we apply the Taylor Expansion to the term inside the exponential, this becomes:

$$k_i(T) = A\mathrm{e}^{\frac{-E_a}{R}\left(\frac{1}{T_0}-\frac{T-T_0}{T_0^2}+\frac{(T-T_0)^2}{T_0^3}-\ldots\right)}, \tag{4.4}$$

assuming that $|T - T_0| \ll T_0$ and rearranging terms yields:

$$k_i(T) \approx \mathrm{e}^{k_{i,0}} \mathrm{e}^{\alpha_{i,T} \Delta T}, \tag{4.5}$$

where $k_{i,0} = -\log(A) E_a / RT$ and $\alpha_{i,T} = E_a / RT_0^2$. However, because the Arrhenius equation assumes the sign of the dependence (i.e., higher temperatures lead to faster rates), and such effects are not known for the other system parameters, we choose the more general formulation detailed in Eq. 4.2.

For each experimental condition, multiple model classes are considered. Each model class consists of a unique combination of transition rate (k_i) dependencies on the experimental variable of interest. This results in a set of model classes for each experimental variable that encompass all possible combinations in which zero, one, two, three, or four of the transition rates are dependent upon that variable.

The concentrations for each element of the oligo-plasmid system evolves over time, and can be described by the coupled set of Ordinary Differential Equations (ODE) shown in Eq. 4.6 - 4.9.

$$\frac{d[C]}{dt} = -k_o[C] + k_c[O] \tag{4.6}$$

$$\frac{d[O]}{dt} = k_o[C] - k_c[O] - k_b[O][U] + k_u[B] \tag{4.7}$$

$$\frac{d[B]}{dt} = k_b[O][U] - k_u[B] \tag{4.8}$$

$$\frac{d[U]}{dt} = -k_b[O][U] + k_u[B] \tag{4.9}$$

where $[X]$ indicates the concentration of X. Because the total numbers of plasmids and oligos are both assumed to be constant, these four ODEs can be simplified with two algebraic constraints:

$$[O]_T = [B] + [U], \tag{4.10}$$

$$[P]_T = [C] + [O] + [B]. \tag{4.11}$$

where $[O]_T$ is the total oligo concentration and $[P]_T$ is the total plasmid concentration. This allows for the oligo-plasmid system to be described using two ODEs instead of four, which are reduced forms of Eq. 4.6 and 4.8 in terms of only [C] and [B]:

$$\frac{d[C]}{dt} = -k_o[C] + k_c([P]_T - [C] - [B]) \tag{4.12}$$

$$\frac{d[B]}{dt} = k_b([P]_T - [C] - [B])([O]_T - [B]) - k_u[B] \tag{4.13}$$

Initially, the concentrations of DNA plasmids in the C and O conformations (closed and open structures) are in equilibrium, with no oligo probes available for binding. Using the total plasmid concentration $[P] = [C] + [O] + [B] = 21.1$ nM, known experimentally, the initial state vector is expressed as:

$$[[C]_0,\ [B]_0]] = \left[\frac{k_c}{k_o + k_c}[P]_T,\ 0\right]. \tag{4.14}$$

In each experiment, oligos are added to the solution at time t_0, and we estimate a 300 second delay to account for the time required to prepare and load the samples before imaging.

For each experimental data set (denoted by $e = 1, \ldots, N_{Exp}$) and time point (denoted by t_i for $i = 1, \ldots, N_t$), the fraction of oligos that are bound to plasmids was measured as $\mu_S(t_i, e)$, and the standard error of this measurement was estimated as $\sigma_{SEM}(t_i, e)$, both using 10 independent measurements scaled per 100 wells. To find the model that most accurately captures these experimental data, we assume that the measurement noise is Gaussian and independent for each time point such that the likelihood to observe the observed fraction of bound oligos in a given experiment is

$$L_e = C_e \prod_{i=1}^{N_t} \exp\left(\frac{-(x(t_i; \mathbf{\Lambda}, e) - \mu_S(t_i, e))^2}{2\sigma_{SEM}^2(t_i, e)}\right), \tag{4.15}$$

where $x(t_i; \mathbf{\Lambda}, e)$ is the model-predicted fraction of bound oligos in experiment e at time point t_i and using parameters $\mathbf{\Lambda}$, and C_e is a normalization constant that does not depend on the parameters. Since all experiments are independently conducted, the likelihood to capture all experiments at once is the product of the individual likelihood values:

$$L = \prod_{e=1}^{N_{Exp}} L_e = \prod_{e=1}^{N_{Exp}} C_e \prod_{i=1}^{N_t} \exp\left(\frac{-(x(t_i;\mathbf{\Lambda},e) - \mu_S(t_i,e))^2}{2\sigma_{SEM}^2(t_i,e)}\right). \tag{4.16}$$

The total loss function for the parameter optimization is the natural logarithm of this likelihood function:

$$J_{fit} = -\log(L) = C + \sum_{e=1}^{N_{Exp}} \sum_{i=1}^{N_t} \frac{(x(t_i;\mathbf{\Lambda},e) - \mu_S(t_i,e))^2}{2\sigma_{SEM}^2(t_i,e)}, \tag{4.17}$$

and is equivalent to the sum of squared scaled residuals (SSR) between the data and each model. We minimize this loss function using MATLAB's genetic algorithm followed by MATLAB's *fminsearch* to determine the parameter set, $\mathbf{\Lambda}$, which minimizes the SSR. We constrain $k_{i,0}$ such that -18<$k_{i,0}$<18 and $\alpha_{i,y}$ so that -1000<$\alpha_{i,y}$<1000, which allows kinetic parameters to range over more than 14 orders of magnitude, but reduces numerical issues associated with unrealistically large or small rate parameters.

Parameter searches for all models are conducted iteratively from many different initial parameter guesses, allowing all models to exchange parameters during the initial stages of the genetic algorithm; all model searches are continued until all fits have converged for every individual model. The models that minimize the loss function, J_{fit}, most accurately reproduce the data and are kept for further testing.

As an additional measure to focus on models that capture the experimental data without overfitting, we perform leave-one-out cross-validation to determine which models best predict data that is held out during fitting. Finally, we calculate the overall cross-validation error, $\langle J_{tot}\rangle_{CV}$:

$$\langle J_{tot}\rangle_{CV} = \frac{1}{N_{CV}} \sum_{i=1}^{N_{CV}} J_{tot}, \tag{4.18}$$

where J_{tot} is the total error for each individual cross-validation model based upon its fit error, J_{fit}, and prediction error, J_{pred}. Models with the lowest overall cross-validation error, $\langle J_{tot}\rangle_{CV}$, demonstrate the most accurate combinations of fits and predictions and are retained in the best classes of models, given the available training data.

Combined Model

Using the model classes that best fit the experimental data to each variable independently, a combined model is constructed that best fits the data while considering all variables at once. By considering all variables simultaneously during the fitting, this combined model more accurately determines which variables are most important for each conformation transition. To construct the general models, we use combinations of the two to four model classes with the smallest $\langle J_{tot} \rangle_{CV}$ for each individual experimental variable (temperature, ionic strength, and supercoiling). Applying the Metropolis-Hastings (MH) algorithm [5, 61] to these model classes estimates the parameter set, $\mathbf{\Lambda_{MH}}$, that results in the best model with multiple dependencies, and also quantifies the uncertainty associated with each parameter. For the proposal function in the MH algorithm, we use a Gaussian distribution with a covariance matrix equal to the inverse of the Fisher Information Matrix (FIM) [154]. This is computed using finite differences at the best fit parameters and known experimental conditions, though many sampling strategies exist when performing an MCMC search [62, 155]. This choice was motivated by the Cramér-Rao lower bound (CRLB) [156], which states that the inverse of the FIM gives a lower bound on the variance of unbiased maximum likelihood estimations (MLE). It was reasoned that the posterior distribution, given the data set, would be broadly similar in shape to the spread of MLE estimates from many independent data sets. We use the $\mathbf{\Lambda}$ values obtained from the previously determined models as the initial sampling point for the MH algorithm. Also, we use a *thin rate* of 50 in the MH search, meaning we only record every 50^{th} accepted sample and discard all others, decreasing storage requirements for the large chains generated during this process. Mixing was confirmed using Geweke diagnostics [157] and integrated autocorrelation times (IACT) in MATLAB's MCMCSTAT package.

4.2 Results and Discussion

To identify the optimal model that captures the dynamics of the interacting oligo-plasmid system as a function of T, $[S]$, and σ, each condition is first modeled individually. As discussed above, the overall cross-validation (CV) error is the metric used to determine the best individual models. The individual results for each variable are subsequently employed to identify the best model capable of fitting all environmental conditions simultaneously. The MH algorithm is then used to estimate the parameters and their uncertainties for this combined model.

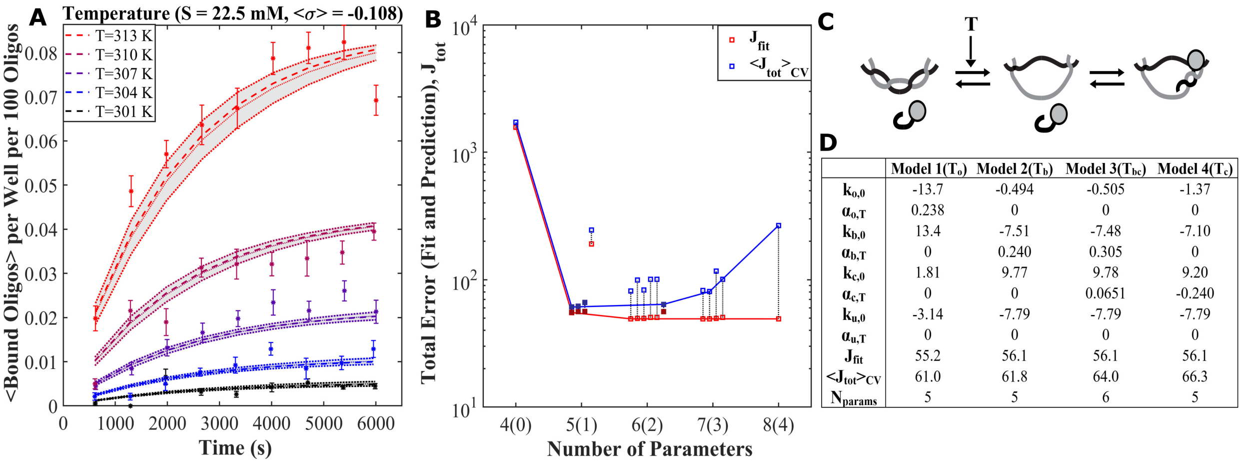
4.2.1 Temperature

Five experimental data sets for the number of binding events per 100 oligonucleotide molecules are considered when modeling T effects, each at a different temperature, T (301, 304, 307, 310, 313)K. Each of these experiments was conducted at the reference ionic strength, $[S]$, of 22.5 mM and the reference supercoiling, $\langle\sigma\rangle$, of -0.108. As a result, the full functional form of the rate equation (Eq. 4.2) for T-dependent transitions is reduced to:

$$k_i(T) = e^{k_{i,0}+\alpha_{i,T}(T-T_0)}. \tag{4.19}$$

The experimental data and fit results for DNA oligo-plasmid binding as a function of time for the best model class for T-dependent binding are shown in Fig. 4.2A, along with shaded regions that demonstrate the variation in model fits resulting from cross-validation. It is clear in this figure that DNA plasmid-probe binding increases as T is increased.

The J_{fit} and $\langle J_{tot}\rangle_{CV}$ for T-dependent binding for all model classes are shown in Fig. 4.2B. When each model has reached convergence, we verify that the total error when the model is fit to all data ($J_{tot} = J_{fit}$) is always lower than the total error ($J_{tot} = J_{fit} + J_{pred}$) when that model is fit to only four data sets and then used to predict the final data set. This can be seen by comparing the corresponding red and blue symbols for each model class in Fig. 4.2B. We also verify that for every possible pair of nested models and every combination of fitting data, the more complex model always produces a better fit than the less complex model. Additionally, for a given number



	Model 1(T_o)	Model 2(T_b)	Model 3(T_{bc})	Model 4(T_c)
$k_{o,0}$	-13.7	-0.494	-0.505	-1.37
$\alpha_{o,T}$	0.238	0	0	0
$k_{b,0}$	13.4	-7.51	-7.48	-7.10
$\alpha_{b,T}$	0	0.240	0.305	0
$k_{c,0}$	1.81	9.77	9.78	9.20
$\alpha_{c,T}$	0	0	0.0651	-0.240
$k_{u,0}$	-3.14	-7.79	-7.79	-7.79
$\alpha_{u,T}$	0	0	0	0
J_{fit}	55.2	56.1	56.1	56.1
$<J_{tot}>_{CV}$	61.0	61.8	64.0	66.3
N_{params}	5	5	6	5

Figure 4.2: Results for Temperature-Dependent Binding. A) Experimental data and fit with Model 1_{T_o} of binding per 100 oligos, averaged over 10 different videos versus time at different T. Data are shown with symbols and error bars that are the standard error of the mean (SEM) of binding per 100 oligos, averaged over 10 videos. The thick dashed lines correspond to the best simultaneous fit to all data. Shaded regions correspond to the mean plus or minus one standard deviation from the mean using cross validation, leaving out one time trace at a time. B) Model Errors (SSR) for the best models at each combination of either zero, one, two, three, or four T-dependent k_i. The total number of parameters ($k_{i,0} + \alpha_{i,y}$) is shown on the x-axis, along with the corresponding number of T-dependent $\alpha_{i,y}$ in parentheses. The red symbols correspond to J_{fit} for the models that best fit all of the data. Blue symbols indicate the $\langle J_{tot}\rangle_{CV}$ for all data when the models are fit to all combinations of four out of the five trajectories. Filled symbols indicate the models in D. Models are clustered by numbers of parameters and ranked left to right according to fit errors. C) Schematic of the model with the best $\langle J_{tot}\rangle_{CV}$, which has one T-dependent rate, $\{k_o\}$. D) Parameters, error values, and the number of parameters for four models with the best $\langle J_{tot}\rangle_{CV}$. Model $1(T_o)$ is depicted in C. Standard deviations for the $\langle J_{tot}\rangle_{CV}$ for Models $1(T_o)$ through $4(T_c)$ are 9.02, 8.76, 9.73, and 12.75, respectively.

of parameters, there is an optimal model class that exhibits a better overall CV error, as shown by the blue symbols in Fig. 4.2B. The blue line visualizes the best overall CV error as a function of the number of parameters and demonstrates that the optimal number of T-dependent parameters is one. Below this optimal number of parameters, the best overall CV error is higher because the models have too few parameters to adequately fit the data. With more than this optimal number of parameters, the model is overfitting the data and is unable to predict the held out data.

Figure 4.2C illustrates the model selected by cross validation for T-dependent binding, and Fig. 4.2D shows the corresponding parameter values for this model, Model $1(T_o)$, which has one T-dependent transition rate: k_o. Whether temperature enhances or inhibits a transition is indicated by the signs of $\alpha_{o,T}$, $\alpha_{c,T}$, and $\alpha_{b,T}$. A positive $\alpha_{i,T}$ corresponds to an increase in k_i with temperature, and a negative $\alpha_{i,T}$ corresponds to a decrease in k_i with temperature. All four models in Fig. 4.2D have very similar J_{fit} and $\langle J_{tot} \rangle_{CV}$, with the strongest similarities between Model $1(T_o)$ and Model $2(T_b)$. Due to these similarities, as well as an equal number of parameter in Model $1(T_o)$ and Model $2(T_b)$, it is difficult to clearly identify which of these two models is definitively better. This difficulty is magnified by the results presented in Table 4.1, which show the mean parameter values plus or minus one standard deviation from cross-validation for the top four models.

For Model $1(T_o)$, substantial variation is present in the CV results for $k_{b,0}$, $k_{c,0}$, and $k_{u,0}$. For $k_{c,0}$ and $k_{u,0}$, most of this variation is a result of the CV model in which the data at the highest temperature, 313 K, is left out. This CV model has a higher total error,J_{tot}, than the others also, which is mostly due to the poor predictive capability of this model for that specific data set compared to the other CV models. However, for the CV model with a very different $k_{b,0}$ value from the others is the one in which the data at the same temperature as the reference temperature, 310 K, is left out. For Model $2(T_b)$, there is less variability in the CV results for many of the parameters, except there is substantially variability in $k_{o,0}$. Once again, this variability is due to the CV model in which the data at the highest temperature, 313 K, is left out, which once again also has a higher total error,J_{tot}, than the others due to the poor predictive capability of this model for that data set as compared to the other CV models.

Model $1(T_o)$ demonstrates that increasing T has a slight stimulatory effect on the opening transition rate (k_o), as can be seen by the positive $\alpha_{o,T}$. The stimulatory effect of T on DNA unwinding is not surprising: as temperature increases, there is more thermal energy that can drive DNA denaturation. The small magnitude of this stimulatory effect may be due to the second unwinding site (Site 2) present on the pUC19 plasmid, which unwinds at higher energies (caused by increasing T or σ), and has been shown experimentally to take some of the energy used by Site 1 for unwinding [3, 158]. According to Model $1(T_o)$, the increase in plasmid-probe interactions seen in the experimental data is due to an enhancement in DNA opening, increasing the site's exposure to its complementary oligo probe. These results have important biological implications:

Table 4.1: Cross-Validation (CV) Results for Temperature-Dependent Binding. Mean parameters from cross-validation, error values, and the number of parameters for the four models with the best $\langle J_{tot} \rangle_{CV}$, shown in Fig. 4.2. CV results are presented as the mean plus or minus one standard deviation from the mean.

	Model 1(T_o)	Model 2(T_b)	Model 3(T_{bc})	Model 4(T_c)
$k_{o,0}$	-13.3 ± 0.829	-0.603 ± 0.439	-0.956 ± 3.72	-0.777 ± 2.61
$\alpha_{o,T}$	0.236 ± 0.00947	0	0	0
$k_{b,0}$	10.2 ± 7.46	-8.20 ± 0.286	-10.2 ± 4.27	-9.16 ± 4.85
$\alpha_{b,T}$	0	0.238 ± 0.00999	0.152 ± 0.131	0
$k_{c,0}$	3.75 ± 4.14	9.30 ± 0.424	8.72 ± 1.74	8.12 ± 1.70
$\alpha_{c,T}$	0	0	- 0.115 ± 0.131	-0.247 ± 0.0240
$k_{u,0}$	-3.96 ± 2.25	-7.81 ± 0.162	-7.82 ± 0.158	-7.82 ± 0.158
$\alpha_{u,T}$	0	0	0	0
J_{fit}	55.2	56.1	56.1	56.1
$<J_{tot}>_{CV}$	61.0 ± 9.02	61.8 ± 8.76	64.0 ± 9.73	66.3 ± 12.8
N_{params}	5	5	6	5

at physiological temperatures (37°C) that match the higher experimental T presented here, binding sites may be more available to interact with important cellular machinery.

In contrast, Model $2(T_b)$ demonstrates a stimulatory effect on the probe-plasmid binding rate (k_b). This effect follows from reaction kinetics, where more thermal energy drives molecular diffusion, increasing the probability for the reactants to collide and react. Interestingly in this model, while the closing rate is independent of T ($\alpha_{c,T} = 0$), the overall rate is significantly higher than the opening, binding, and unbinding rates (i.e., $k_{c,0} >> k_{o,0}$, $k_{b,0}$, and $k_{u,0}$). For biological systems, a very high closing rate is problematic: interactions with DNA sites that close too quickly would be difficult as these sites would only be available for interaction for very short time periods. Since neither opening nor closing are thermally-driven according to this model, higher temperatures would be required for cells to function properly as it would sufficiently increase binding enough to overcome this steric hindrance. Model $3(T_{bc})$ in Fig. 4.2D has two T dependencies: k_b and k_c. Here, T has a stimulatory effect on both the probe plasmid binding rate (k_b) and the closing rate of the plasmid (k_c). Despite an additional T-dependency, the magnitudes of all of the transition rates are surprisingly similar between this model and Model $2(T_b)$, suggesting redundancy in the use of model parameters when considering temperature data alone. Additional data to quantify the combined effects of supercoiling and ionic strength in addition to temperature could further refine the $\langle J_{tot} \rangle_{CV}$ and J_{fit} values, and allow for a clearer distinction between models for this system, as we will explore in the subsequent sections.

4.2.2 Supercoiling

The flexibility of DNA structural changes is heavily controlled by the under- or over-twisting of the DNA double-helix, known as *supercoiling*. Supercoiling is quantified through the linking number *Lk*, which is defined as the sum of the twist *Tw* (the number of times a DNA double-helix turns about itself) and the writhe *Wr* (where one part of the DNA double-helix crosses over another) [3,159]. As the DNA double-helical structure is naturally twisted, torsional strain (or supercoiling) applied to a DNA molecule is quantified via the linking number difference $\Delta Lk = Lk - Lk_0$, where *Lk* is the linking number of a supercoiled DNA molecule, and Lk_0 is the linking number of the same molecule with no supercoiling. Unstressed DNA (i.e., with no supercoiling) has a natural Tw rate of 10.4 bp/turn under physiological conditions, and no Wr. As the amount of torsional stress for DNA with a particular ΔLk decreases with increasing size, the superhelical density $\sigma = \Delta Lk \times 10.4/N$ is a better measure of the superhelical stress on a DNA molecule. Here, *N* is the number of base pairs of the molecule, and 10.4 is the twist rate of a relaxed DNA molecule.

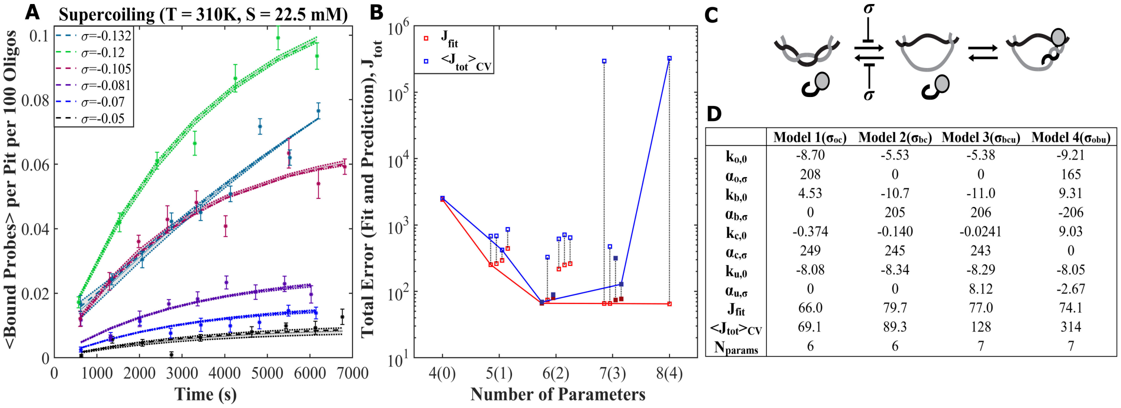
Torsional stress in DNA created by supercoiling provides energy to drive structural transitions, such as allowing the standard B-DNA double helix to become unwound DNA, Z-DNA, or DNA cruciforms [159]. Cells can modify supercoiling through DNA-protein interactions, such as through the use of topoisomerases and gyrases. Additionally, transcribing RNA polymerases can generate sufficient transient supercoils to drive structural transitions hundreds to thousands of base pairs away, which may serve functional purposes. This is the case with the far upstream element (FUSE) region associated with the c-myc oncogene in humans. Here, transient supercoils created by the RNA polymerases cause the FUSE region to unwind. FUSE binding proteins (FBP) can then bind to the unwound FUSE region, stimulating c-myc transcription, which in turn generates more supercoils, creating a positive feedback loop that ends when FBP Interacting Repressor (FIR) proteins bind to the FUSE region and return the c-myc transcription to baseline levels [160].

To observe plasmid-probe interaction dynamics as a function of supercoiling, experimental data sets counting the number of binding events per 100 probes over time were collected at different σ: $\langle\sigma\rangle$ =-0.05, -0.07, -0.081, -0.105, -0.12, and -0.132. As the pUC19 plasmids in each experiment had a Gaussian-distribution of σ, the mean $\langle\sigma\rangle$ is used; see below for further details. Each experiment was conducted at a reference temperature of $T = 310$ K and the reference of 22.5 mM ionic strength. As only $\langle\sigma\rangle$ varied for these experiments, the full functional form of the transition rate equation (Eq. 4.2) for supercoiling-dependent transitions is reduced to:

$$k_i(\sigma) = e^{k_{i,0}+\alpha_{i,\sigma}(\sigma-\sigma_0)}. \tag{4.20}$$

Supercoiled pUC19 plasmids used for experiments were grown and extracted from *Escherichia coli* cells using standard plasmid purification methods with a Qiafilter Plasmid Midi Kit (Qiagen, Toronto, ON, Canada) [3, 161]. Supercoiling itself was quantified by running the plasmids through an agarose gel, which separates plasmids of different σ due to slight differences in their size [3]. Plasmids were supercoiled to different negative superhelicities by reacting them with topoisomerase IB in the presence of varying quantities of ethidium bromide. Typically, plasmids extracted and reacted in such a way have a Gaussian-distribution of supercoiling; as in the previous work, $\langle\sigma\rangle$ is the mean σ obtained from a Gaussian fit of the plasmids in the agarose gel. To determine the σ of each plasmid within a particular sample, a control sample of pUC19 is reacted with either a topoisomerase IA or IB enzyme. This removes most supercoiling present in the sample, such that $\langle\sigma_0\rangle = 0$. Using this as a reference, and a ladder composed of plasmids with known values of σ, the $\langle\sigma\rangle$ of a sample may be determined. A negative superhelical density, σ, corresponds to underwound DNA. Note that structural tension in supercoiled plasmids depends upon the magnitude of supercoiling, and not its sign.

The experimental data and best fits for plasmid-probe binding averaged over 10 different videos per 100 probes as a function of time for several $\langle\sigma\rangle$ are shown in Fig. 4.3A. Shaded regions indicate the variation in model fits resulting from cross-validation. As seen in Fig. 4.3A, plasmid-probe binding is nonlinear with $\langle\sigma\rangle$; binding increases with the magnitude of $\langle\sigma\rangle$ up to $\langle\sigma\rangle = -0.120$, then decreases at $\langle\sigma\rangle = -0.132$. At this supercoiling level, a secondary unwinding site (Site 2) begins to unwind, absorbing some of the torsional energy used by Site 1 [158, 162]. With less Site 1 unwinding, probe-plasmid binding is reduced at these supercoiling levels.



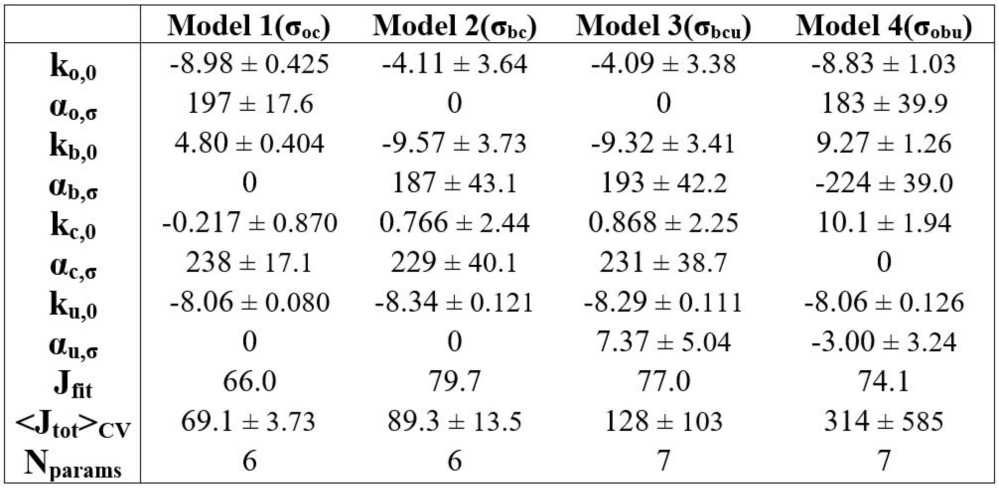
	Model 1(σ_{oc})	Model 2(σ_{bc})	Model 3(σ_{bcu})	Model 4(σ_{obu})
$k_{o,0}$	-8.70	-5.53	-5.38	-9.21
$\alpha_{o,\sigma}$	208	0	0	165
$k_{b,0}$	4.53	-10.7	-11.0	9.31
$\alpha_{b,\sigma}$	0	205	206	-206
$k_{c,0}$	-0.374	-0.140	-0.0241	9.03
$\alpha_{c,\sigma}$	249	245	243	0
$k_{u,0}$	-8.08	-8.34	-8.29	-8.05
$\alpha_{u,\sigma}$	0	0	8.12	-2.67
J_{fit}	66.0	79.7	77.0	74.1
$\langle J_{tot}\rangle_{CV}$	69.1	89.3	128	314
N_{params}	6	6	7	7

Figure 4.3: Results for Supercoiling-Dependent Binding. See Fig. 4.2 for symbol descriptions. A) Experimental data and fit with Model 1_{σ_oc} of binding averaged over 10 different videos per 100 probes versus time for a range of $\langle\sigma\rangle$ (supercoiling). B) Model Errors (SSR) for the best models at each combination of either zero, one, two, three, or four $\langle\sigma\rangle$-dependent k_i. The total number of parameters ($k_{i,0} + \alpha_{i,\sigma}$) is shown on the x-axis, along with the corresponding number of $\langle\sigma\rangle$-dependent $\alpha_{i,\sigma}$ in parentheses. C) Schematic of the model with the best $\langle J_{tot}\rangle_{CV}$, which has two $\langle\sigma\rangle$-dependent parameters, $\{k_o, k_c\}$. D) Parameters, error values, and the number of parameters for four models with the best $\langle J_{tot}\rangle_{CV}$. Model $1(\sigma_{oc})$ is depicted in C. Standard deviations for the $\langle J_{tot}\rangle_{CV}$ for Models $1(\sigma_{oc})$ through $4(\sigma_{obu})$ are 3.73, 13.5, 103, and 585, respectively.

The J_{fit} (total error when a model is fit to all data) and $\langle J_{tot}\rangle_{CV}$ (average error when a model is fit to five data sets and used to predict the final one) for σ-dependent binding for all model classes are shown in Fig. 4.3B. Unlike the T-variant data in which one dependent transition rate was optimal, Fig. 4.3B suggests that two σ-dependent transition rates are optimal to fit and predict the data. It is important to note that because supercoiling increases as $\langle\sigma\rangle$ becomes more negative, a negative (positive) $\alpha_{i,\sigma}$ in Fig. 4.3D results in an increase (decrease) in k_i.

The best model (Model $1(\sigma_{oc})$) for σ-dependent binding is shown in Fig. 4.3C, with the corresponding fitting parameter values shown in Fig. 4.3D. Model $1(\sigma_{oc})$ in Fig. 4.3D has two σ-

Table 4.2: Cross-Validation Results for Supercoiling-Dependent Binding. Mean parameters from cross-validation, error values, and the number of parameters for the four models with the best $\langle J_{tot} \rangle_{CV}$, shown in Fig. 4.3. CV results are presented as the mean plus or minus one standard deviation from the mean.



	Model 1(σ_{oc})	**Model 2(σ_{bc})**	**Model 3(σ_{bcu})**	**Model 4(σ_{obu})**
$k_{o,0}$	-8.98 ± 0.425	-4.11 ± 3.64	-4.09 ± 3.38	-8.83 ± 1.03
$\alpha_{o,\sigma}$	197 ± 17.6	0	0	183 ± 39.9
$k_{b,0}$	4.80 ± 0.404	-9.57 ± 3.73	-9.32 ± 3.41	9.27 ± 1.26
$\alpha_{b,\sigma}$	0	187 ± 43.1	193 ± 42.2	-224 ± 39.0
$k_{c,0}$	-0.217 ± 0.870	0.766 ± 2.44	0.868 ± 2.25	10.1 ± 1.94
$\alpha_{c,\sigma}$	238 ± 17.1	229 ± 40.1	231 ± 38.7	0
$k_{u,0}$	-8.06 ± 0.080	-8.34 ± 0.121	-8.29 ± 0.111	-8.06 ± 0.126
$\alpha_{u,\sigma}$	0	0	7.37 ± 5.04	-3.00 ± 3.24
J_{fit}	66.0	79.7	77.0	74.1
$<J_{tot}>_{CV}$	69.1 ± 3.73	89.3 ± 13.5	128 ± 103	314 ± 585
N_{params}	6	6	7	7

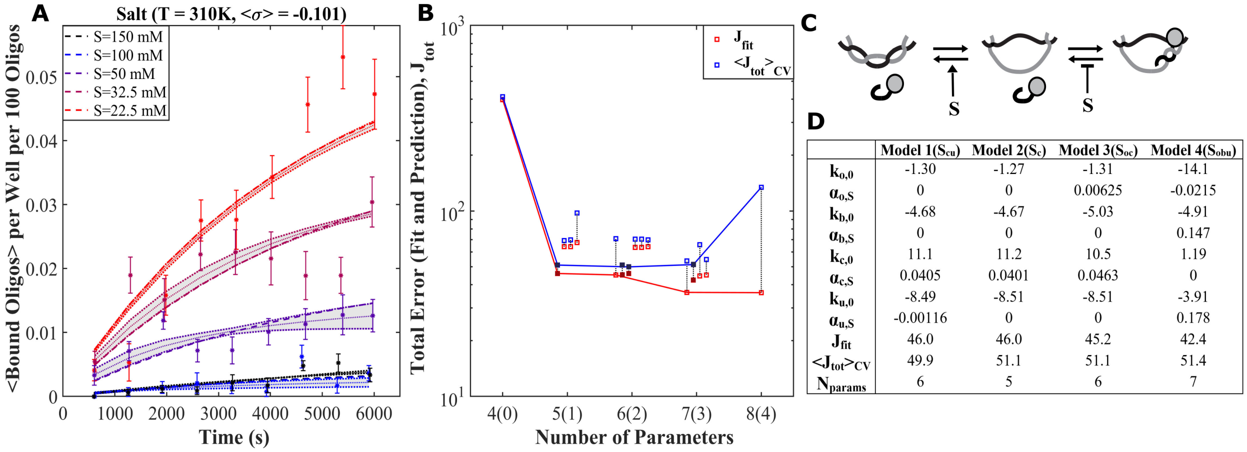
dependent k_i and two independent of σ; only opening and closing depend on σ. Comparing Model $1(\sigma_{oc})$ with the others, one sees that J_{fit} and $\langle J_{tot} \rangle_{CV}$ for Model $1(\sigma_{oc})$ are noticeably lower than those of the other models, indicating that the optimal model is significantly better than the others at both fitting the included data and predicting the left out data set. In fact, Model $1(\sigma_{oc})$ can fit the left out data set with the highest supercoiling level quite well, even though binding at Site 1 becomes nonlinear at this torsional stress. As with T, higher supercoiling provides more torsional stress and therefore more structural energy. With enough structural energy, Site 2 begins to unwind, simultaneously stealing some of the energy Site 1 uses to unwind, causing a decrease in binding. For Model $1(\sigma_{oc})$, J_{tot} is slightly higher (73.4 and 74.3 versus $66 - 67$) for the CV models that leave out the data sets at the highest and lowest σ values,$\langle \sigma \rangle = -0.132$ and $\langle \sigma \rangle = -0.05$, respectively, due to higher predictive error for these two data sets as compared to the others, possibly related to the nonlinearity in binding behavior. Looking at Table 4.2, there is large variability in $k_{c,0}$ for Models $1(\sigma_{oc})$ through $3(\sigma_{bcu})$, with a standard deviation at least 2 times larger than the mean. This is a partially the result of $k_{c,0}$ changing signs when the data at the highest σ value, in the nonlinear region, is left out during cross validation.

That Model $1(\sigma_{oc})$ predicts that supercoiling has an inhibitory effect on unwinding and rewinding and no effect on binding or unbinding is not surprising: supercoiling is inherently a structural phenomenon in the plasmid alone. Though supercoiling enables binding with the unwinding site, it should neither have any effect on increasing the collisions that drive oligo probe interactions, nor on forcing the probe to unbind. What is interesting here is that the model predicts that both opening and closing are inhibited by increased supercoiling. While this seems counterintuitive, if the rates of opening and closing are compared, it is evident that closing is far more inhibited with increasing supercoiling, as shown by $\alpha_{o,\sigma} < \alpha_{c,\sigma}$. Thus, at higher supercoiling, closing is far more impeded than opening, permitting the unwinding site to be open for longer times than it is closed, as is the case for lower supercoiling ($\frac{k_o(\sigma_1)}{k_c(\sigma_1)} < \frac{k_o(\sigma_2)}{k_c(\sigma_2)}$ for $|\sigma_1| < |\sigma_2|$).

The dependence of these structural rate changes on supercoiling demonstrates how important torsional tension is to DNA structural transitions. This is important to consider physiologically, where torsional tension in cellular DNA is controlled by topoisomerase enzymes which can relax, or further supercoil, DNA. Judging by the dependence of k_o and k_c on σ presented here, improper topoisomerase function could have a negative effect on DNA structural transitions, increasing or decreasing DNA unwinding compared to normal levels within a cell and affecting subsequent interactions. Cancer and other genetic diseases can result in humans when DNA σ is misregulated [17, 18]; the model presented here reinforces the idea that this may be due to a supercoil-drive change in structural transitions, subsequently affecting DNA interactions.

4.2.3 Ionic Concentration

To evaluate the effect of ionic concentration on plasmid-probe binding, experiments counting the number of binding events averaged over 10 different videos per 100 probes were collected at a range of ionic concentrations: 22.5 mM, 32.5 mM, 50 mM, 100 mM, and 150 mM. These concentrations were calculated using NaCl concentrations of 10 mM, 20 mM, 37.5 mM, 87.5 mM, and 137.5 mM, and considering the ionic contributions of the 12.5 mM Tris and 25 mM HEPES present at a pH of 8.0. Each experiment was conducted at the reference temperature of $T = 310$ K using pUC19 plasmids with $\langle\sigma\rangle = -0.101$. Note that this value is less supercoiled than the reference supercoiling level. Given this, the reduced functional form of the transition rate equation



	Model 1(S_{cu})	Model 2(S_c)	Model 3(S_{oc})	Model 4(S_{obu})
$k_{o,0}$	-1.30	-1.27	-1.31	-14.1
$\alpha_{o,S}$	0	0	0.00625	-0.0215
$k_{b,0}$	-4.68	-4.67	-5.03	-4.91
$\alpha_{b,S}$	0	0	0	0.147
$k_{c,0}$	11.1	11.2	10.5	1.19
$\alpha_{c,S}$	0.0405	0.0401	0.0463	0
$k_{u,0}$	-8.49	-8.51	-8.51	-3.91
$\alpha_{u,S}$	-0.00116	0	0	0.178
J_{fit}	46.0	46.0	45.2	42.4
$<J_{tot}>_{CV}$	49.9	51.1	51.1	51.4
N_{params}	6	5	6	7

Figure 4.4: Results for Salt-Dependent Binding. See Fig. 4.2 for symbol descriptions. A) Experimental data and fit with Model $1(S_{cu})$ of binding averaged over 10 different videos per 100 probes versus time for a range of $[S]$. B) Model Errors (SSR) for the best models at each combination of either zero, one, two, three, or four $[S]$-dependent k_i. The total number of parameters ($k_{i,0}+\alpha_{i,S}$) is shown on the x-axis, along with the corresponding number of $[S]$-dependent parameters in parentheses. C) Schematic of the model with the best $\langle J_{tot}\rangle_{CV}$, which has two $[S]$-dependent parameters, $\{k_c, k_u\}$. D) Parameters, error values, and the number of parameters for four models with the best $\langle J_{tot}\rangle_{CV}$. Model $1(S_{cu})$ is depicted in C. Standard deviations for the $\langle J_{tot}\rangle_{CV}$ for Models $1(\sigma_{oc})$ through $4(\sigma_{obu})$ are 16.2, 6.82, 6.25, and 9.76, respectively.

(Eq. 4.2) for salt-dependent transition rates is:

$$k_i([S]) = e^{k_{i,0}+\alpha_{i,S}([S]-[S_0])}. \tag{4.21}$$

Similar to the T and σ-dependent results, the fit of Model $1(S_{cu})$ as well as the experimental data for plasmid-probe binding as a function of time and salt concentration $[S]$ are shown in Fig. 4.4A. In contrast to the binding vs. T and σ where plasmid-probe binding increases with the experimental variable, here, binding decreases as $[S]$ is increased, resulting in the lowest plasmid-probe binding at the highest ionic strength. Also worth noting is the higher prediction error for the data at the lowest salt concentration (22.5 mM) for all of the models shown in Fig. 4.4, which may be related to the higher SEM for the experimental data at this concentration, as seen by the error bars in Fig. 4.4A.

The J_{fit} and $\langle J_{tot}\rangle_{CV}$ for $[S]$ binding for all model classes are shown in Fig. 4.4B. While Fig. 4.4B-D show that a model with two $[S]$-dependent k_i optimally fits the data, with the Model $1(S_{cu})$ being the best model for $[S]$-dependent binding, all models have very similar J_{fit} and $\langle J_{tot}\rangle_{CV}$ values. Despite the differences in salt-dependencies between Models $1(S_{cu})$, $2(S_c)$, and $3(S_{oc})$, the magnitudes of all four transition rates are incredibly similar due to the small magnitudes of $\alpha_{u,S}$ in Model $1(S_{cu})$ and $\alpha_{o,S}$ in Model $3(S_{oc})$, with Model $1(S_{cu})$ being the least likely model to accurately model the system, as no unbinding of oligo probes was ever observed experimentally. Additionally, the results from cross-validation, presented in Table 4.3, demonstrate much more variability in parameters and $\langle J_{tot}\rangle_{CV}$ for Model $1(S_{cu})$ compared to both Model $2(S_c)$ and Model $3(S_{oc})$. Given this, the only meaningful dependency on salt concentration demonstrated in Models $1(S_{cu})$, $2(S_c)$, and $3(S_{oc})$ is k_c, which is the rate of transitioning from the open (unwound) conformation the wound (closed) conformation. The enhancement of Site 1 rewinding with increasing $[S]$ is possibly due to cations screening the negatively-charged DNA backbone. The negative charges on each DNA strand repulse those on the other, decreasing the amount of thermal or torsional energy needed for site unwinding. As ionic strength increases, more dissolved cations are attracted to the negatively charged DNA backbone, reducing the electrostatic repulsion, and thus enhancing rewinding.

Table 4.3: Cross-Validation Results for Salt-Dependent Binding. Mean parameters from cross-validation, error values, and the number of parameters for the four models with the best $\langle J_{tot}\rangle_{CV}$, shown in Fig. 4.4. CV results are presented as the mean plus or minus one standard deviation from the mean.

	Model 1(S_{cu})	**Model 2(S_c)**	**Model 3(S_{oc})**	**Model 4(S_{obu})**
$k_{o,0}$	-6.48 ± 6.98	-1.36 ± 0.261	-1.26 ± 0.666	-14.0 ± 0.215
$\alpha_{o,S}$	0	0	0.00264 ± 0.00521	-0.0213 ± 0.00268
$k_{b,0}$	-2.57 ± 2.95	-4.74 ± 0.310	-4.84 ± 0.633	-3.31 ± 3.40
$\alpha_{b,S}$	0	0	0	0.100 ± 0.0963
$k_{c,0}$	9.01 ± 2.55	11.2 ± 0.119	10.8 ± 0.466	2.05 ± 1.82
$\alpha_{c,S}$	0.0756 ± 0.042	0.0401 ± 0.00237	0.0436 ± 0.00642	0
$k_{u,0}$	-6.61 ± 2.86	-8.62 ± 0.832	-8.62 ± 0.833	-3.26 ± 2.02
$\alpha_{u,S}$	-0.0260 ± 0.0279	0	0	0.132 ± 0.0944
J_{fit}	46.0	46.0	45.2	42.4
$<J_{tot}>_{CV}$	49.9 ± 16.2	51.1 ± 6.82	51.1 ± 6.25	51.4 ± 9.76
N_{params}	6	5	6	7

Model $4(S_{obu})$, however, has substantially different transition rates than the other three models. Model $4(S_{obu})$ agrees with DNA structural theory, where DNA secondary structures are assumed to have an inverse dependence for unwinding on $[S]$ [159]. As highlighted above, this inverse dependence arises from salt screening of the electrostatic repulsion present in the negatively-charged DNA phosphate backbone. In this model, however, the effect of this salt screening on the repulsion causes a subsequent decrease in unwinding, as opposed to enhancement of rewinding. As discussed above for Model $1(S_{cu})$, the fourth model has a salt-dependent probe unbinding, which was never observed experimentally, rendering this model less probable as an accurate model for this system.

Due to the similarities in the J_{fit} and $\langle J_{tot}\rangle_{CV}$ values, it is difficult to clearly identify which model is definitively better than the other without additional experimental data, although Model $2(S_c)$ seems to be the most likely based on both the results presented in Fig. 4.4, the physical understanding of the system, and the cross-validation results presented in Table 4.3. The combined model described in the next section, however, was developed to test multiple combinations of dependencies on all experimental variables ($[S]$, T, and $\langle\sigma_0\rangle$) to elicit which model best fits the data.

All of the models in Fig. 4.4D indicate that DNA structural changes, either through an increase in rewinding, or a decrease in unwinding, are dependent upon ionic concentration. Consequently, important cellular processes that rely on open binding sites, such as replication or transcription, could possibly be enhanced (or impeded) by changes in salt concentration. This idea finds support in the presence of ion channels within cells that maintain a homeostatic ionic concentration within cells [163].

4.2.4 Temperature, Ionic Concentration, and Supercoiling

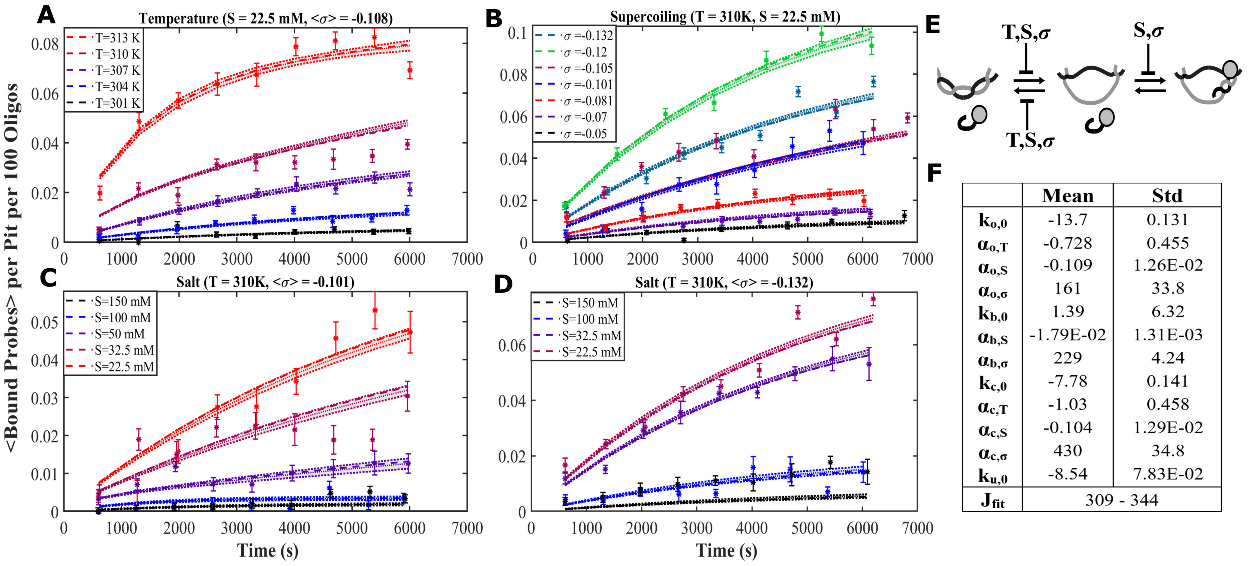
To determine the optimal model with the least number of parameters that can fit all of the data simultaneously, all three of the experimental variables (T, σ, and $[S]$) must be taken into consideration. In addition to the experimental data sets independently modeled above, new data sets are considered here for plasmids with $\langle\sigma\rangle = -0.132$ for three ionic strengths: 32.5 mM, 100 mM, and 150 mM. These experiments were conducted at the reference temperature of $T =$

310 K. It should be noted that at this high supercoiling, the binding behavior to Site 1 is nonlinear as a result of Site 2 unwinding.

When considering T, $[S]$, and σ simultaneously, evaluating all of the combinatorially large number of possible models is time intensive and computationally expensive. To reduce computation time, we used the previous models obtained by fitting each variable independently, reducing the total number of models evaluated. Specifically, we combined the dependencies from the top model class for each individual experimental variable, ignoring unlikely models. For the models obtained for T and $[S]$, only the rate dependencies from the top two models are considered when fitting all data simultaneously, due to their similar fit errors. For σ, the transition rate dependencies from the top four models in Fig. 4.3D are considered as the nonlinear behavior of the plasmid-probe interactions create a large variation in model results. We also included the most general model, where all structural transitions are dependent upon each variable (T, $[S]$, and σ), in case any behavior cannot be adequately captured when modeling each variable independently, resulting in a total of 45 models. In contrast to the cross-validation approach used for the individual experimental conditions, a Metropolis-Hastings (MH) search is performed after fitting the data to estimate parameter uncertainties because of the prohibitive number of possible cross-validation data sets.

Applying this method to the data, six models are identified with similarly low (and overlapping) fit error distributions. To determine which of these best captures the physical behavior of the plasmid-probe system, we compare the parameter dependencies identified in these models. For all six models, the fit error for the salt and supercoiling data is high, while the fit error for the temperature data is relatively low. As both the salt and supercoiling data were obtained with plasmids at $\langle\sigma\rangle = -0.132$, nonlinear binding behavior dominates due to Site 2 opening, potentially accounting for the higher error. For all models, the value of $\alpha_{u,S}$ switches signs between positive and negative during the MH search, behavior indicating that this transition is independent of $[S]$, and can therefore be ignored. Additionally, $\alpha_{b,T}$, $\alpha_{u,T}$, and $\alpha_{b,\sigma}$ are zero in several of the best-fitting models; these rates are also ignored.

Eliminating these rate dependencies results in a final model with reduced parameter dependencies, shown in Fig. 4.5. The shaded regions in Fig. 4.5 correspond to the mean plus or minus one standard deviation from the mean for 25 randomly-sampled parameter sets from the MH search. This reduced model includes temperature dependencies from Model $4(T_o)$ and Model $4(T_c)$ in Fig. 4.2D, the supercoiling (σ) dependencies from Model $1(\sigma_{oc})$ and Model $2(\sigma_{bc})$ in Fig. 4.3D, and the salt dependencies from Models $2(S_c)$, and $3(S_{oc})$ in Fig. 4.4D, along with the additional $[S]$-dependency of the rate of binding.



	Mean	Std
$k_{o,0}$	-13.7	0.131
$\alpha_{o,T}$	-0.728	0.455
$\alpha_{o,S}$	-0.109	1.26E-02
$\alpha_{o,\sigma}$	161	33.8
$k_{b,0}$	1.39	6.32
$\alpha_{b,S}$	-1.79E-02	1.31E-03
$\alpha_{b,\sigma}$	229	4.24
$k_{c,0}$	-7.78	0.141
$\alpha_{c,T}$	-1.03	0.458
$\alpha_{c,S}$	-0.104	1.29E-02
$\alpha_{c,\sigma}$	430	34.8
$k_{u,0}$	-8.54	7.83E-02
J_{fit}	309 - 344	

Figure 4.5: Results for Binding with Dependencies on Temperature, Ionic Concentration, and Supercoiling. A) Experimental data and model fit of binding versus time at different temperatures (A), superhelical densities (B), and ionic strengths (C,D). Data are shown with symbols and the error bars represent the SEM of binding for 10 averaged videos. The thick dashed lines correspond to the best simultaneous fit to all data. Shaded regions correspond to the mean plus or minus one standard deviation from 25 randomly-sampled parameter sets from the MH search results. (E) Schematic and (F) parameters of the final, reduced model.

In the previous Ionic Concentration section (Sec. 4.2.3), when considering DNA-oligo binding solely as a function of salt concentration, it was difficult to identify the best model due to similarities in the fit error and overall cross-validation error between all of the models, although Models $1(S_{cu})$ and $4(S_{obu})$ are the least likely models to accurately model the system, as no unbinding of oligo probes was ever observed experimentally. Given this difficulty, along with the large SEM of the data at the two lowest salt concentrations in comparison to the other data, as seen in Fig. 4.5 and Fig. 4.4A, it is unsurprising that the combined results had different salt dependencies than when evaluated independently. This is especially true given the inclusion of an additional salt data when fitting all data simultaneously. It is also possible that these three transition dependencies upon $[S]$ were only observable when considering the interplay between all three system variables at once: T, σ, and $[S]$. As such, modeling DNA structural and interaction kinetics may only truly capture some behavior in a system when all variables are considered simultaneously, as opposed to one variable independently, as performed in previous sections.

This final model also allows us to compare the impact of each parameter (temperature, ionic strength, and supercoiling) on each possible structural transition of the system: unwinding, rewinding, binding, and unbinding. It should be noted that the magnitudes of each $\alpha_{i,y}$ are not comparable between different y values. Of particular note are the dependencies of three structural transitions on both $[S]$ and σ, with unbinding completely independent of all variables. The independence of binding upon T is surprising: according to reaction kinetics, interactions between the oligo and plasmid should increase exponentially with increasing T. It is likely that within this relatively small range of T, the structural dynamics are far more sensitive than reaction kinetics to temperature changes. This has implications for cellular processes that require molecules, such as proteins, to bind to DNA, as in DNA replication or transcription. Sub-optimal replication or transcription in cells at lower than physiological temperatures may be due in greater part to interference in DNA unwinding than changes in diffusion.

The dependence of three structural transitions (unwinding, rewinding, and binding) on σ and $[S]$ demonstrates how both structural dynamics and reaction kinetics rely on these variables. Of particular note is the weaker dependence of binding upon these variables: the effect of supercoiling or ionic strength on structural dynamics dominates over oligo binding. This indicates that for DNA-binding processes, it is the sensitivity of DNA unwinding itself to σ and $[S]$ that is important, and not their effect on DNA-DNA interactions. It's well-known that a change in supercoiling can have an impact on gene expression in cells [160]; this combined model supports this hypothesis.

Additionally, the presence of cellular mechanisms to control ionic strength, such as ion channels in either the cellular or nuclear membranes, demonstrate the importance of salt concentration to molecules and processes in cells, such as with DNA interactions. Also, temperature control within cells is quite difficult, usually requiring their growth in thermally stable environments at their species' preferred conditions. As such, there is very little cell-based thermal regulation of DNA and DNA processes. It follows from this that temperature would have as much of an influence on DNA structure and interactions as either supercoiling or ionic strength. Of these two variables, supercoiling is even more tightly regulated via numerous classes and species of proteins, from topoisomerases to histones. The importance of exact supercoiling control, on both a global and local level, is evident in our results by its strong effect on the kinetic rates of all the DNA processes studied here, with the exception of DNA unbinding. These results are also pertinent to gene editing techniques, such as CRISPR-Cas9, which employ a single-stranded nucleotide that binds to a target DNA region. It is likely that the DNA structure of the target region may play an important role in the success of these gene editing processes, and that the role of supercoiling, salt conditions, and temperature cannot be discounted.

How each of the three variables affects structural dynamics is also important for the subsequent interactions. While both unwinding and rewinding are negatively-dependent upon T and σ, rewinding is even more sensitive to both of these mechanisms than unwinding. Thus, at higher T or σ, once Site 1 unwinds, it is more likely to stay unwound than at lower values of these variables, making it more available to bind other molecules, such as the oligo probe used here. This is further support for the idea that the DNA structure is a dominant force in reaction kinetics, as opposed to the diffusive process between the plasmid site and the oligo. This has implications for the study of DNA interactions with other molecules, as DNA structural dynamics clearly plays an important role in the interactions or reactions. Furthermore, the negative T-dependence of this general model disagrees with the optimal model (Model $1(T_o)$) determined in Sec. 4.2.1, which indicates an increase in the rate of unwinding as temperature increases. While ionic strength affects both unwinding and rewinding, the rates of each seem to decrease nearly identically with increasing $[S]$, while for the reaction kinetics, only the binding interaction is negatively affected. Thus, while both the structural and reaction kinetics are affected by ionic strength, an overall decrease in interactions is observed.

What is also surprising is the effect of supercoiling on DNA binding. Supercoiling in this system is only applied to the pUC19 plasmid, in theory affecting only its structure. As posited in the Supercoiling section (4.2.2), it's possible that further supercoiling of pUC19 not only promotes DNA-DNA interactions through increased DNA unwinding, but also by making the unwinding site more available for binding. As supercoiling increases in a plasmid, the DNA molecule's writhe also increases substantially. This restricts its possible conformations to those with multiple crossovers of the DNA double-helix. It is possible that a toroid-like structure could be created, as was posited when pUC19 molecules were crowded [4]. This would also be advantageous in living cells as it would provide additional control over DNA interactions; by reducing the overall supercoiling, it could hide interaction sites within the DNA structure, whereas increasing overall supercoiling would increase interactions.

Although we have a model that fits the data very well and while some of the sign dependencies in this model are understood, it is likely there are some interactions and physical mechanisms that are not entirely understood or accounted for in this model when all data is considered simultaneously, particularly related to the interconnectivity between the experimental variables, T, σ, and $[S]$.

This is demonstrated in the exploration of parameter space using the MH algorithm, as seen in a small subset of the parameter results in Fig. 4.7. In Fig. 4.7, *sloppiness* can be seen, with parameters that are highly uncertain in some directions, despite being well-constrained in others [112, 113]. Given this, the system parameters may only be partially identifiable with this model. Also, additional analyses are needed to understand the bimodal probability distribution of J_{fit} from the MH search of parameter space, seen in Fig. 4.6. It is also possible that issues related to the bias-variance trade-off for this model have been encountered [164–166] and a revised modeling approach is needed [167]. This is further suggested by the substantially higher fit error obtained

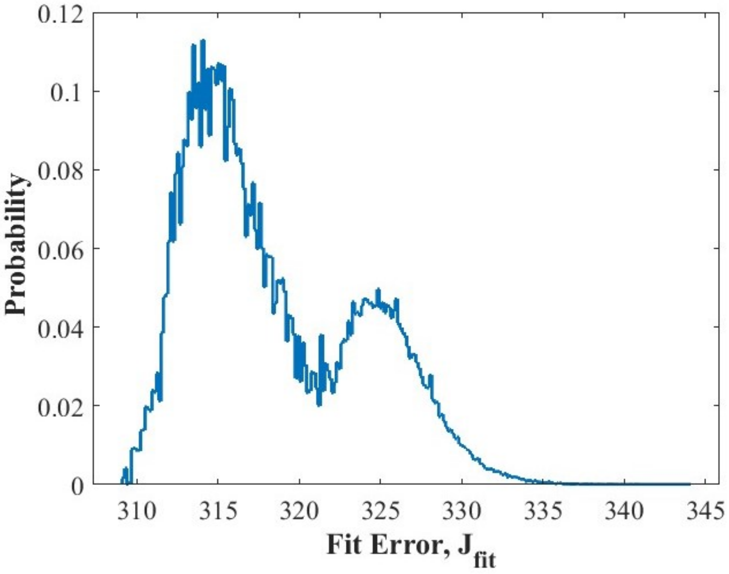


Figure 4.6: Probability distribution of the MLE fit error, J_{fit}, from the Metropolis-Hastings (MH) search for the final, reduced model for binding with dependencies on temperature, ionic concentration, and supercoiling.

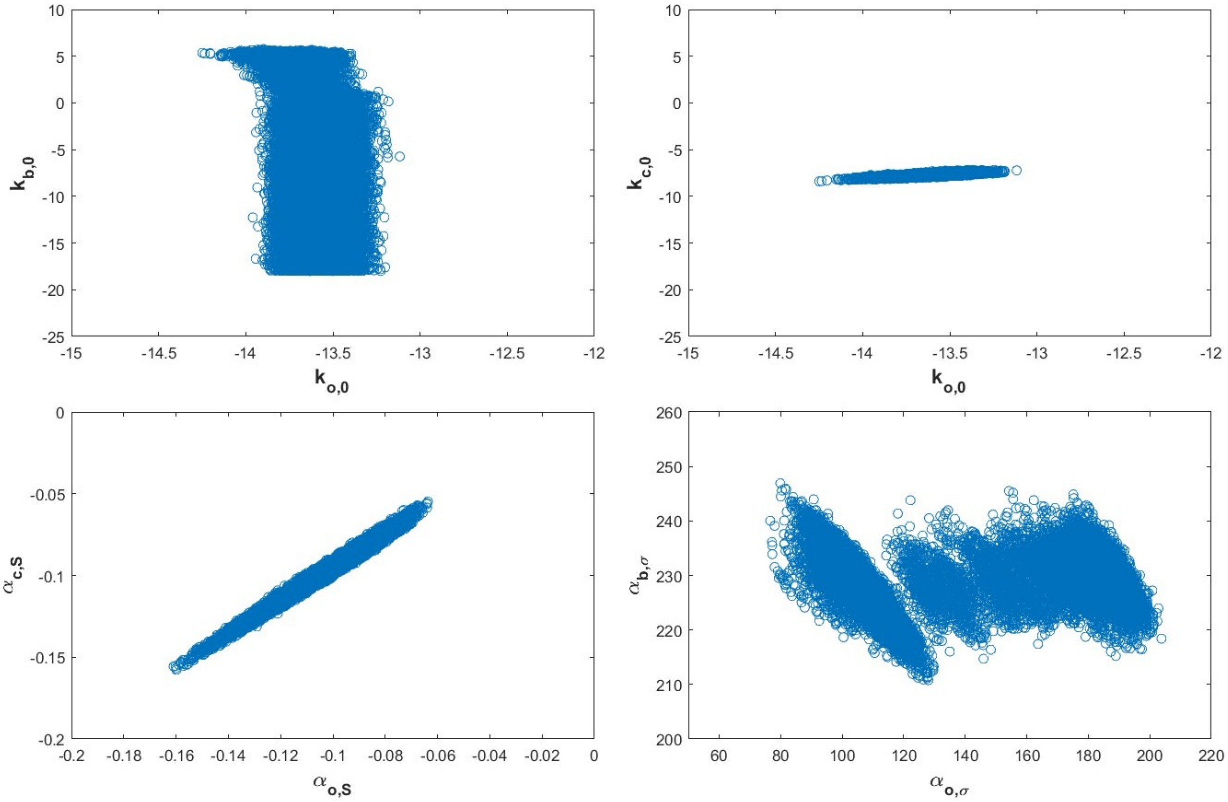


Figure 4.7: Results from the Metropolis-Hastings exploration of parameter space for the final, reduced model for binding with combined dependencies on temperature, ionic concentration, and supercoiling.

with an intuitive model to all of the data as compared to the fit error for the final, reduced model presented here. Unlike the final, reduced model, the intuitive model incorporates sign restrictions consistent with the physical understanding of the system; the intuitive model results in a fit error, J_{fit}, of 547. Since the final, reduced model is, in essence, a nested, reduced version of the intuitive model, once sign restrictions based on the intuitive understanding are systematically removed, the fit error improves and the model dependencies become similar to those present the final, reduced model. Directly related, when sign restrictions are incorporated in the final, reduced model based on intuition of the physical mechanisms, the fit gets worse. Given this, further investigation and potentially more data are needed to resolve these issues related to interpretation of the final model [167].

4.3 Conclusions

Determining how DNA structure and subsequent interactions depend upon temperature, supercoiling, and ionic strength has important ramifications for cellular processes, such as DNA transcription or replication. By applying a computational approach to model the interactions between an oligo and known unwinding site in a pUC19 plasmid, we obtain interesting results on how each variable affects the structural and reaction kinetics. The computational approach used here demonstrates that considering how each variable affects the kinetics independently provides different results than when each variable is considered simultaneously, potentially as a result of synergistic impact of the experimental conditions on DNA conformational changes and binding behavior.

Our combined model indicates that for the DNA-DNA interactions studied in this system, temperature, supercoiling and ionic strength play a much larger role in regulating structural dynamics than reaction kinetics. As such, how these variables affect DNA structural dynamics may be more important than previously thought for DNA interaction processes, such as for gene editing techniques. Surprisingly, our numerical models show that supercoiling may have an effect on DNA-DNA interactions, potentially through the creation of another structure, as can occur under crowded conditions. Our numerical models thus reinforce the idea that supercoiling has an important effect on DNA structure and subsequent interactions, with consequences for plasmid replication and gene transcription. Further investigation is needed in the form of both additional experimental data and potentially a revised modeling approach, however, to improve the understanding of the interplay between these experimental conditions on the structural and binding dynamics of the system as well as better elucidate the model as related to the intuitive physical understanding of the system.

Chapter 5

Diversity, Equity, and Inclusion in Undergraduate Chemical and Biological Engineering Curriculum[4]

5.1 Introduction

> "... without diversity we limit the set of life experiences that are applied, and as a result we pay an opportunity cost — a cost in products not built, in designs not considered, in constraints not understood, and in processes not invented."*Wulf (2001)*

What are diversity, equity, and inclusion? According to Dr. Jessica Rush Leeker at the University of Colorado at Boulder (CU Boulder), "In the simplest terms, diversity refers to the presence of differences in any sense. Equity refers to fairness. And inclusion refers to having resources and opportunity." [169]. Improving diversity in engineering and fostering inclusive professional identities is crucial for more inclusive engineering designs as well as for society as a whole, as concisely summarized in the opening quote by Wulf (2001) [170]. When there is a limited representation of women and historically excluded groups in engineering, including groups that are racially-minoritized, there is also a limited view of where there is a need for engineers, both for design of new or improved products and development or creation of new processes. From Mays (2022), "Cognitive diversity generates more and better ideas, and identity diversity fuels cognitive diversity. But regarding equity and inclusion, diversity does not work without them" [171].

Engineering students should be aware of the connection and intersections between these concepts and the engineering career, which is why it is important for diversity and inclusion to be integrated in required engineering courses and taught by engineering faculty and instructors. This

---

[4] Portions of this work were partially funded as part of an NSF-supported initiative. In addition, some of the topics presented in this chapter, particularly the previous work done, which is discussed in Sec. 5.4, were presented at 2021 Collaborative Network for Engineering and Computing Diversity (CoNECD) Conference [168].

is especially critical given exclusionary aspects that continue to endure in the engineering culture, despite some effort in the last few years for improvement [172–174]. These include individualism, meritocracy and the assumption that meritocracy is without bias, the belief that engineering can and should be 'depoliticized' or removed from 'cultural concerns', and the closely related belief that topics, such as gender equality, that are related to 'social' issues should be separate from the technical aspects of engineering [172–175].

In reality, engineering, culture, and public welfare are closely intertwined. The need for engineers is directly dependent upon the needs of society, and an understanding of the needs of society related to engineering applications cannot be separated from the diversity (or lack thereof) of the engineering workforce. According to the 2022 U.S. Census Bureau data for the employed civilian population 16 years and over, females accounted for only 18.97% of bioengineers and biomedical engineers and only 19.97% of chemical engineers [176,177]. Additionally, according to the Digest of Education Statistics, out of all employed 25- to 34-year olds with a Bachelor's degree, 17.4% were engineers. However, the proportion of women compared to men is lower across racial/ethnic backgrounds, with the proportion of Black women being substantially lower than all others for both men and women (see Fig. 5.1).

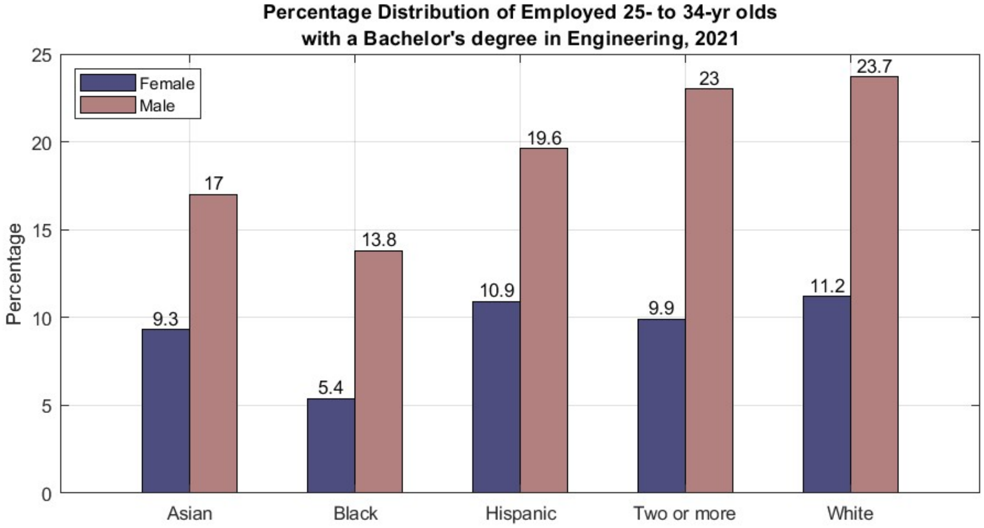


Figure 5.1: Percentage Distribution by race/ethnicity for all employed 25- to 34-year olds with a bachelor's degree in engineering in 2021. Data obtained from Digest of Education Statistics (2021), Table 505.30 [178].

In order to achieve a more diverse and inclusive engineering workforce that is capable of solving the numerous challenges that face society, it is imperative to adapt engineering education to be more inclusive overall, promote an increase in diversity, and focus on improving equity within the field. In addition to the incorporation of DEI into core engineering courses (in both pedagogy and curriculum) in a program, there are many other important topics and concepts related tangentially or less obviously to DEI that can also be incorporated. This can be approached from numerous different angles, some of which include:

- encouraging diverse representation of students, staff, and faculty within the university, college, and program; [179, 180]
- adopting a more inclusive pedagogy that supports the breadth of backgrounds, learning styles, and perspectives as well as making sure course materials are accessible to all students, including persons with disabilities; [181–183]
- fostering a positive, inclusive, respectful classroom environment that considers and accommodates the backgrounds, cultures, experiences, and identities that the students and faculty hold; [174, 182, 184]
- making revisions or adaptations to the curriculum to increase the presence of DEI through the use of modified language, real-world examples, case studies, social and cultural context and sensitivity, and environmental, ethical, and societal considerations; [173, 184]
- ensuring students are aware of current support programs, communities, and resources that are open and available to them at the university, within the engineering college, and within the individual department; [179]
- improving, increasing, or modifying faculty and staff training and educational workshops; [185–187]
- interactive theatre sketches and improvisation activities; [188–191]
- helping students to improve teamwork [184, 192] and learn about implicit biases [193] to assist them in fostering more inclusive professional identities;
- and incorporating community-based and community-engaged engineering projects and problem solving into the program. [194, 195]

Regarding diverse representation, this can be improved both by attracting more women and students from racially-minoritized and historically excluded groups to engineering majors as well as making modifications to improve retention of these students in engineering programs. It is particularly critical that these go hand-in-hand since attracting these students does not make sense if efforts have not been made increase retention of them in the program. These are some of the more obvious areas worth consideration.

5.2 Approach

This chapter addresses the DEI challenge in engineering by focusing on the Chemical and Biological Engineering (CBE) Department at Colorado State University (CSU). The overall goal motivating this work is to attract and graduate a diverse population of students in alignment with the CSU mission of access, and to ensure that graduates of the program are prepared to practice engineering in inclusive and equity-minded ways. The purpose of this chapter is to propose a course of action for the CBE department at CSU to substantially increase the connections between engineering, society, and content related to diversity, equity, and inclusion in the department's curriculum. In pursuit of the goals and purpose of this chapter, the following objectives are addressed:

1. Analyze the current student composition of the CBE department to establish a baseline and identify challenges for improvement;
2. Describe prior work to incorporate DEI into the department curriculum, including lessons learned and potential improvements;
3. Review and synthesize work by other engineering education researchers to lay out proposed changes to the department.

5.3 Background

5.3.1 Walter Scott, Jr. College of Engineering (COE) Demographics

Before considering changes to the undergraduate curriculum or the wide spread adoption of initiatives that may change a program, it is important to understand the current state of a program, including the current demographics and challenges the department faces as well as where there are opportunities for growth and improvement. It is therefore relevant to evaluate the previous and current demographics, graduation rates, and so on of the undergraduate student population in the Walter Scott, Jr. College of Engineering (COE) at Colorado State University. The undergraduate enrollment demographics for the COE at CSU are shown in Fig. 5.2 through 5.5 and Tables 5.1 through 5.5 [196]. Demographics for the undergraduate student population specifically for the CBE department at CSU are shown in Fig. 5.6 through 5.10 [196].[5]. Throughout the rest of this chapter, the abbreviation COE denotes the Walter Scott, Jr. College of Engineering, ENGR denotes Engineering, BME denotes Biomedical Engineering, CBE denotes the Chemical and Biological Engineering department, CIVE denotes the Civil and Environmental Engineering department, ECE denotes the Electrical and Computer Engineering department, EIC denotes Engineering Intra-College, and ME denotes the Mechanical Engineering department.

Table 5.1 and Fig. 5.2 show that the ME department is the largest engineering department, with a higher overall enrollment than the other engineering departments; the smallest department is the CBE department, having the lowest overall enrollment. These numbers, however, do not include students who are in a dual degree program with BME and either CBE, ECE, or ME; in this figure (and as provided by CSU [196]), these students are included in EIC along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education. In Table 5.1 as well as Fig. 5.3 and 5.4, both the numbers corresponding to Fig. 5.2 are shown as well as numbers when the dual degree BME students are removed from EIC and included with their respective corresponding programs (CBE, ECE, or ME).

Looking at both Table 5.1, Fig. 5.2, and Fig. 5.3 the ME department had a maximum number of students in 2015, and decreased overall enrollment until Fall 2020. However, since the Covid-

---

[5] It is assumed that gender refers to biological sex, not gender identity. Enrollment data in engineering at CSU for some additional groups of historically excluded or minoritized groups, such as those who identify as LGBTQ+, has not been found (and may not be collected or tabulated).Despite this, however, LGBTQ+ professionals in STEM fields are an historically excluded and underrepresented group that may experience a wide range of negative workplace and career inequalities [197]. In addition, international students appear to be classified as non-minoritized in the data [196].

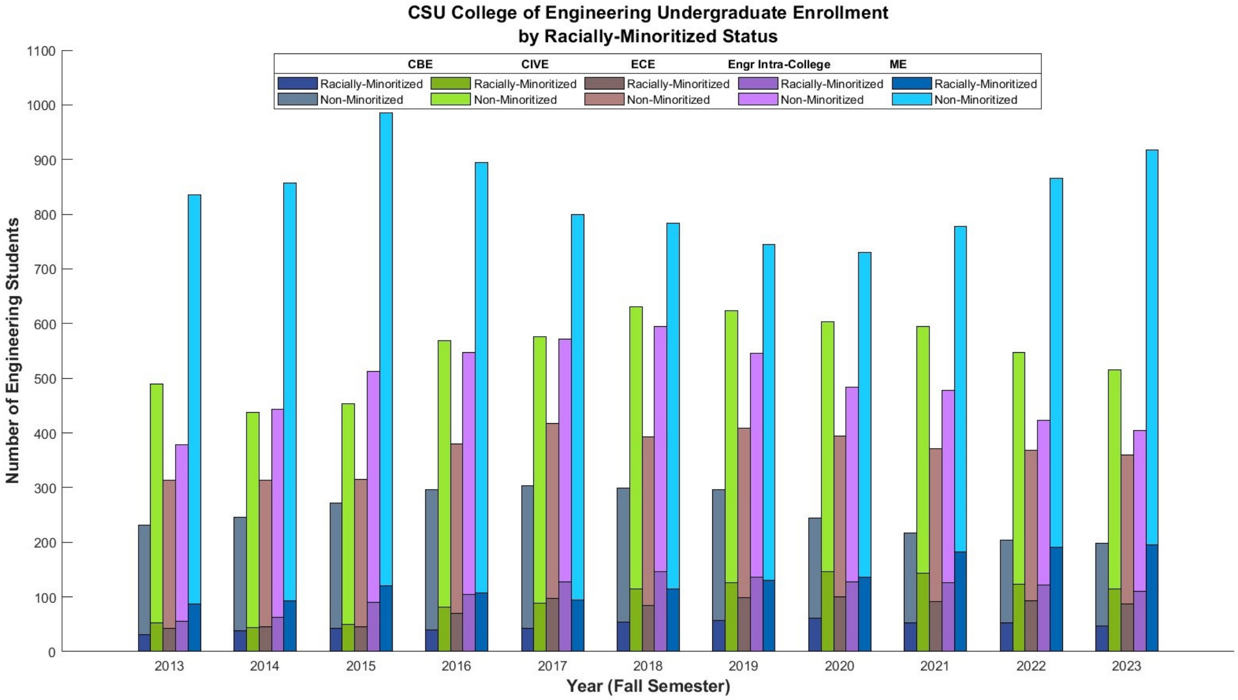
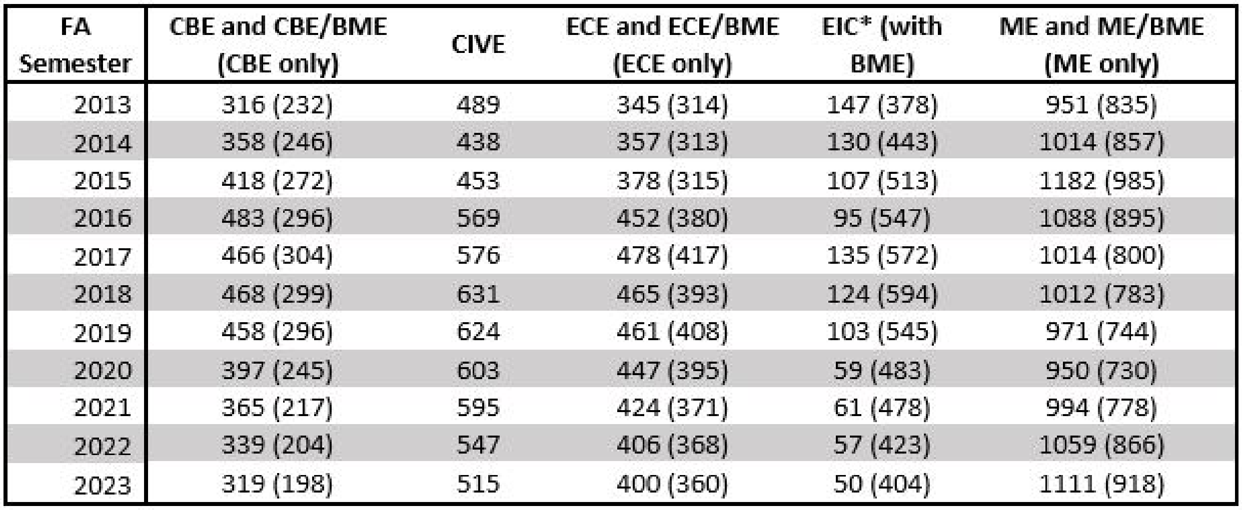


Figure 5.2: Number of racially-minoritized students in the Walter Scott, Jr. College of Engineering (COE) at CSU from 2013 to 2023, by department. Each bar represents the total number of students in the respective department; the total number of students is divided into the number that are racially-minoritized and non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

19 pandemic began impacting enrollment in FA 2020, the only department that has rebounded and seen an increase in enrollment is the ME department, reaching levels just under the Fall 2015 levels in Fall 2023. Enrollment in CBE (both with or without including the dual degree BME students) in Fall 2023 is at levels not seen since Fall 2009 or earlier. In recent years, the ME department has taken steps to focus on: increasing recruitment and retention "from historically marginalized and excluded populations; building an environment where students are culturally competent and have the skills to enter a diverse workforce; improving the campus climate of inclusion; infusing curriculum and educational programs with diversity, equity and inclusion (DEI) content" along with some other areas [199]. However, it is unknown if any of these steps have contributed to the increase in enrollment in recent years.

Table 5.1: Total undergraduate enrollment for the COE at CSU from Fall 2013 to present by Department, including dual degree BME students. For the CBE, ECE, and ME departments, the numbers shown to the left of the parentheses are the *combined* numbers of students enrolled in the specific department and the dual degree BME students, and the numbers in parentheses do not include students enrolled in the dual degree BME program. The * indicates that dual major BME students with CBE, ECE, or ME have been removed from the count for EIC [196].



FA Semester	CBE and CBE/BME (CBE only)	CIVE	ECE and ECE/BME (ECE only)	EIC* (with BME)	ME and ME/BME (ME only)
2013	316 (232)	489	345 (314)	147 (378)	951 (835)
2014	358 (246)	438	357 (313)	130 (443)	1014 (857)
2015	418 (272)	453	378 (315)	107 (513)	1182 (985)
2016	483 (296)	569	452 (380)	95 (547)	1088 (895)
2017	466 (304)	576	478 (417)	135 (572)	1014 (800)
2018	468 (299)	631	465 (393)	124 (594)	1012 (783)
2019	458 (296)	624	461 (408)	103 (545)	971 (744)
2020	397 (245)	603	447 (395)	59 (483)	950 (730)
2021	365 (217)	595	424 (371)	61 (478)	994 (778)
2022	339 (204)	547	406 (368)	57 (423)	1059 (866)
2023	319 (198)	515	400 (360)	50 (404)	1111 (918)

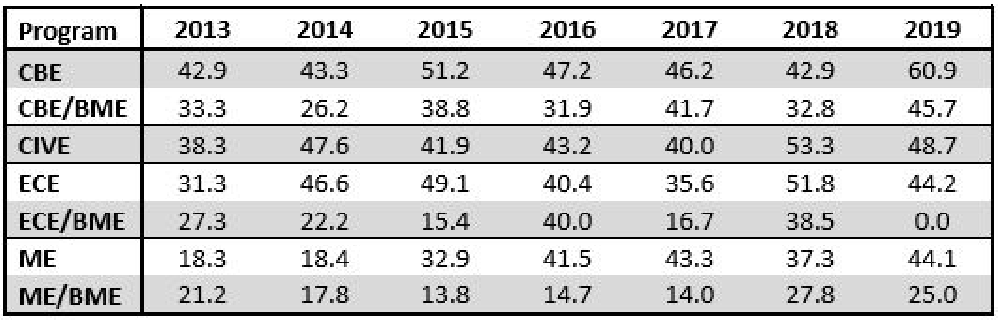
The four- and five-year graduation rates for the primary programs in the COE are shown in Tables 5.2 through 5.7. For each cohort, Tables 5.2 and 5.5 show the graduation rates from any program within the University as a whole, Tables 5.3 and 5.6 show the graduation rates from any program within the College of Engineering, and Tables 5.4 and 5.7 show the graduation rates from the same program in which students began.

Although the four-year graduation rates as shown in Table 5.2 for the most recent graduating cohort (Fall 2019 cohort, which graduated in Spring 2023) show that CBE had the highest percentage (at 60.9%), and the ME department had the lowest percentage (at 44.1%), this figure shows the graduation percentage in four years from any program within the university as a whole, not specifically the College of Engineering or the specific program in which they started. Disregarding the dual degree programs (with BME), which are primarily five-year programs, the four-year graduation rates for students that graduated from any program within the COE (shown in Table 5.3), CIVE and CBE had the highest percentages, which were nearly the same (42.0% and 41.3%, respectively), while ECE had the lowest (with 28.4%). When looking at the four-year graduation rates within the same program, CIVE again had the highest percentage once again and ECE again had the lowest percentage, shown in Table 5.4. Including transfer students in the results increases the four-year graduation rates for the Fall 2019 cohort for all of the primary departments (with graduation rates for ME, ECE, CBE, and CIVE of 39.1%, 30.7%, 40.4%, and 45.3%, respectively) [196].

When looking at the five-year graduation rate for the most recent cohort (Fall 2018, which also graduated in Spring 2023), the CBE department is noticeably lower as compared with other departments and with itself in previous years (as seen in Tables 5.3 through 5.7). It is possible some of this is a result of the 2020 pandemic, which impacted academic years 2020-2021 and 2021-2022 substantially; it is not possible to quantify this with the data available, however. In reality, the full impact of the 2020 Covid-19 pandemic overall will most likely not be understood and quantifiable for many years to come.

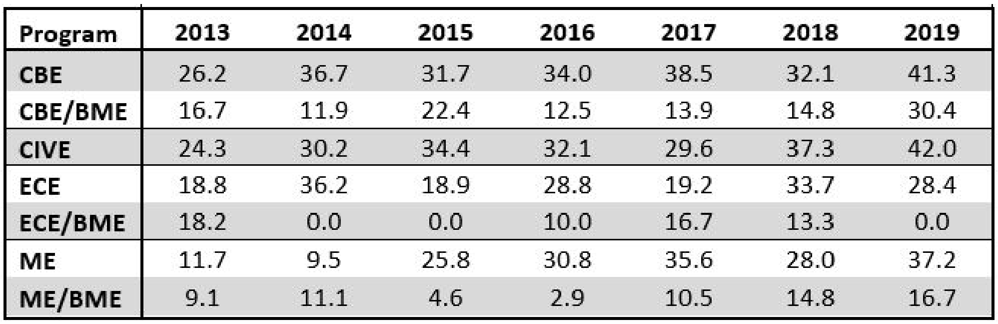
It is also worth noting that these numbers may not be entirely reflective for departments with dual degree BME students since, as previously mentioned, the dual degree BME students are in a different department within the College of Engineering. This will not impact the graduation rates from the University and College of Engineering statistics; however, the graduation rates from the same program (shown in Tables 5.4 and 5.7) will be impacted by this.

Table 5.2: The four-year graduation rate from the University overall By COE Cohort and Department at CSU from Fall 2013 to Fall 2019. For the CBE, ECE, and ME departments, the number of students enrolled in the specific department is shown separately than the dual degree BME students. [196].



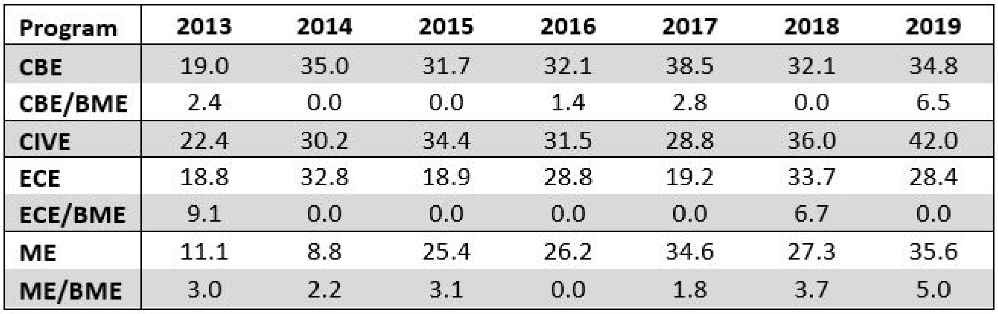
Program	2013	2014	2015	2016	2017	2018	2019
CBE	42.9	43.3	51.2	47.2	46.2	42.9	60.9
CBE/BME	33.3	26.2	38.8	31.9	41.7	32.8	45.7
CIVE	38.3	47.6	41.9	43.2	40.0	53.3	48.7
ECE	31.3	46.6	49.1	40.4	35.6	51.8	44.2
ECE/BME	27.3	22.2	15.4	40.0	16.7	38.5	0.0
ME	18.3	18.4	32.9	41.5	43.3	37.3	44.1
ME/BME	21.2	17.8	13.8	14.7	14.0	27.8	25.0

Table 5.3: The four-year graduation rate from the College of Engineering By COE Cohort and Department at CSU from Fall 2013 to Fall 2019. For the CBE, ECE, and ME departments, the number of students enrolled in the specific department is shown separately than the dual degree BME students. [196].



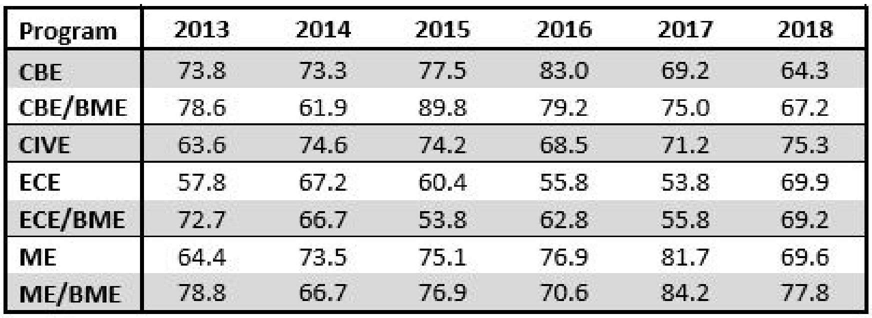
Program	2013	2014	2015	2016	2017	2018	2019
CBE	26.2	36.7	31.7	34.0	38.5	32.1	41.3
CBE/BME	16.7	11.9	22.4	12.5	13.9	14.8	30.4
CIVE	24.3	30.2	34.4	32.1	29.6	37.3	42.0
ECE	18.8	36.2	18.9	28.8	19.2	33.7	28.4
ECE/BME	18.2	0.0	0.0	10.0	16.7	13.3	0.0
ME	11.7	9.5	25.8	30.8	35.6	28.0	37.2
ME/BME	9.1	11.1	4.6	2.9	10.5	14.8	16.7

Table 5.4: The four-year graduation rate from the same program By COE Cohort and Department at CSU from Fall 2013 to Fall 2019. For the CBE, ECE, and ME departments, the number of students enrolled in the specific department is shown separately than the dual degree BME students. [196].



Program	2013	2014	2015	2016	2017	2018	2019
CBE	19.0	35.0	31.7	32.1	38.5	32.1	34.8
CBE/BME	2.4	0.0	0.0	1.4	2.8	0.0	6.5
CIVE	22.4	30.2	34.4	31.5	28.8	36.0	42.0
ECE	18.8	32.8	18.9	28.8	19.2	33.7	28.4
ECE/BME	9.1	0.0	0.0	0.0	0.0	6.7	0.0
ME	11.1	8.8	25.4	26.2	34.6	27.3	35.6
ME/BME	3.0	2.2	3.1	0.0	1.8	3.7	5.0

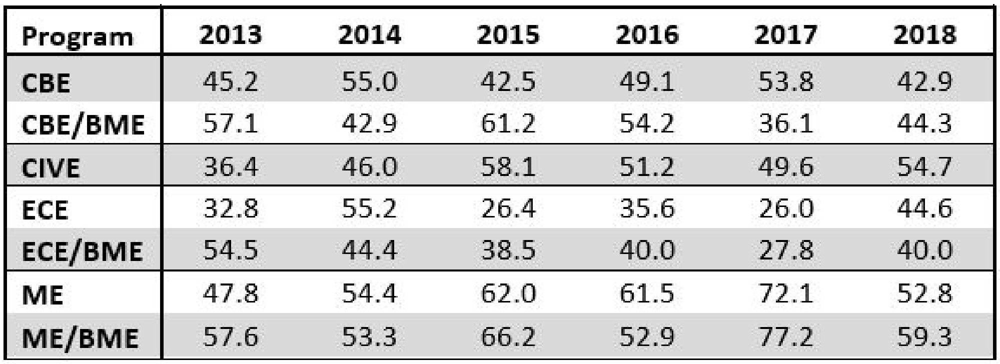
Table 5.5: The five-year graduation rate from the University overall By COE Cohort and Department at CSU from Fall 2013 to Fall 2018. For the CBE, ECE, and ME departments, the number of students enrolled in the specific department is shown separately than the dual degree BME students. [196].



Program	2013	2014	2015	2016	2017	2018
CBE	73.8	73.3	77.5	83.0	69.2	64.3
CBE/BME	78.6	61.9	89.8	79.2	75.0	67.2
CIVE	63.6	74.6	74.2	68.5	71.2	75.3
ECE	57.8	67.2	60.4	55.8	53.8	69.9
ECE/BME	72.7	66.7	53.8	62.8	55.8	69.2
ME	64.4	73.5	75.1	76.9	81.7	69.6
ME/BME	78.8	66.7	76.9	70.6	84.2	77.8

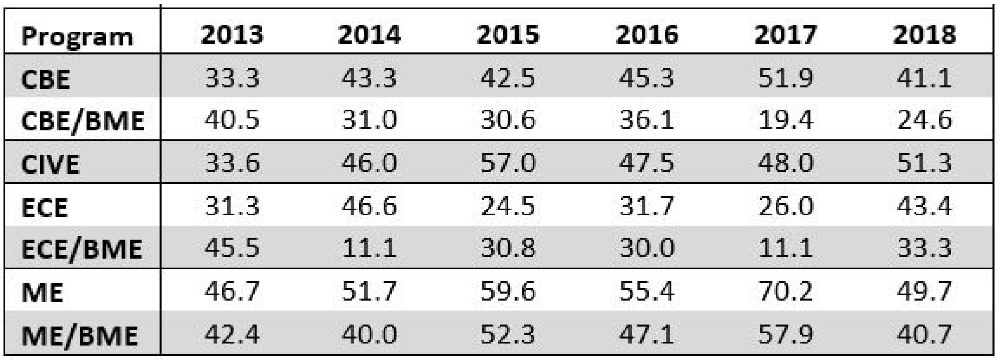
The percentages of racially-minoritized students compared to the departmental undergraduate student population are shown in Fig. 5.4F. The values in parentheses are for the CBE, ECE, and ME departments when the BME dual degree students are not taken into consideration in these percentages. When looking at these values, ME has the lowest percentage of racially-minoritized

Table 5.6: The five-year graduation rate from the College of Engineering By COE Cohort and Department at CSU from Fall 2013 to Fall 2019. For the CBE, ECE, and ME departments, the number of students enrolled in the specific department is shown separately than the dual degree BME students. [196].



Program	2013	2014	2015	2016	2017	2018
CBE	45.2	55.0	42.5	49.1	53.8	42.9
CBE/BME	57.1	42.9	61.2	54.2	36.1	44.3
CIVE	36.4	46.0	58.1	51.2	49.6	54.7
ECE	32.8	55.2	26.4	35.6	26.0	44.6
ECE/BME	54.5	44.4	38.5	40.0	27.8	40.0
ME	47.8	54.4	62.0	61.5	72.1	52.8
ME/BME	57.6	53.3	66.2	52.9	77.2	59.3

Table 5.7: The five-year graduation rate from the same program By COE Cohort and Department at CSU from Fall 2013 to Fall 2019. For the CBE, ECE, and ME departments, the number of students enrolled in the specific department is shown separately than the dual degree BME students. [196].



Program	2013	2014	2015	2016	2017	2018
CBE	33.3	43.3	42.5	45.3	51.9	41.1
CBE/BME	40.5	31.0	30.6	36.1	19.4	24.6
CIVE	33.6	46.0	57.0	47.5	48.0	51.3
ECE	31.3	46.6	24.5	31.7	26.0	43.4
ECE/BME	45.5	11.1	30.8	30.0	11.1	33.3
ME	46.7	51.7	59.6	55.4	70.2	49.7
ME/BME	42.4	40.0	52.3	47.1	57.9	40.7

students, with 21%. Generally, the overall percentage of racially-minoritized undergraduate students has increased over time for most departments and has hovering between $20-30\%$ for at least the last 3-5 years, with only slight fluctuations. When looking at percentage of racially-minoritized students compared to the departmental undergraduate student population, shown in parentheses in Fig. 5.4F, ME has the lowest at 21%. However, when the dual degree BME students are included with their respective corresponding programs (with weighted averages based on the number of students), as shown in Fig. 5.4 and Table 5.1, the undergraduate enrollment of participating departments (CBE, ECE, and ME) increases considerably, as does the percentage of racially-minoritized students within the department when compared to the percentage only based on non-dual degree students for both CBE and ME for nearly every semester. For CBE/BME,

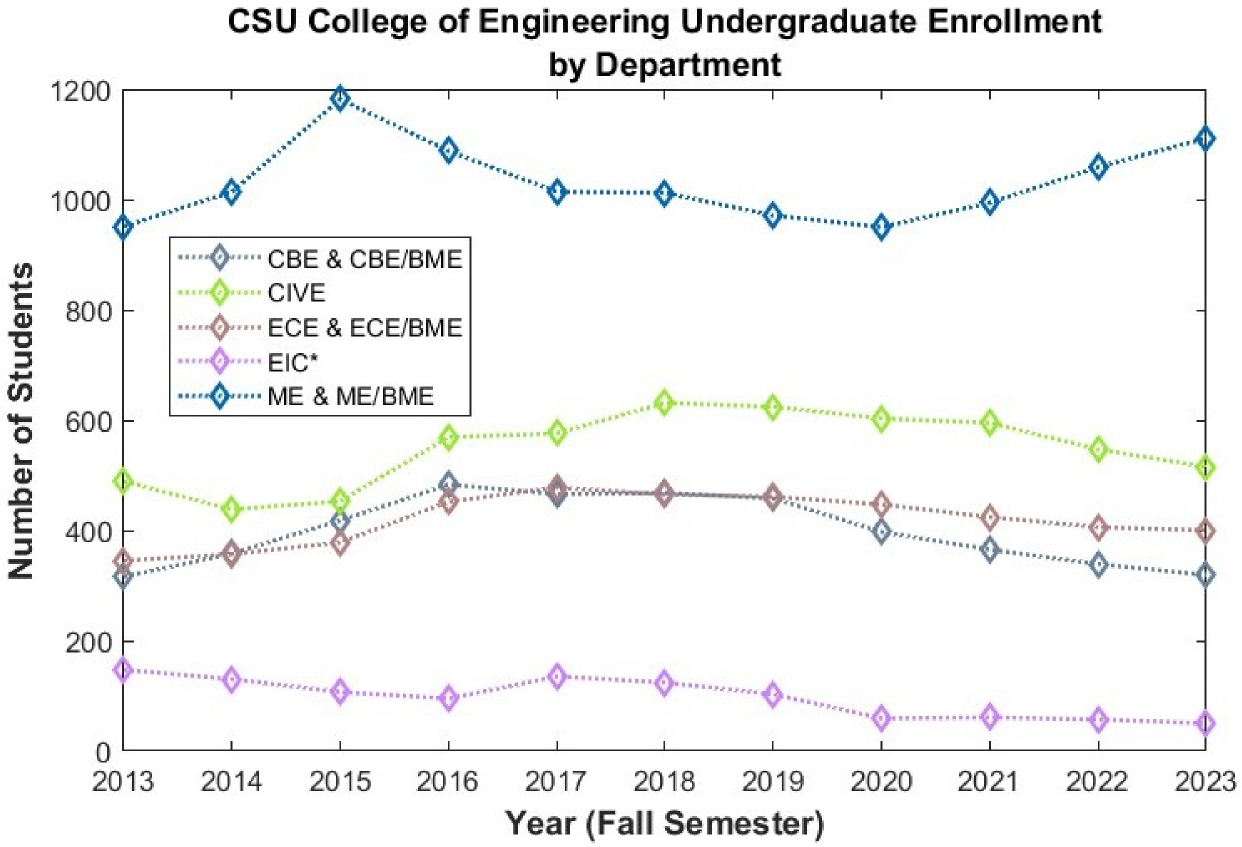
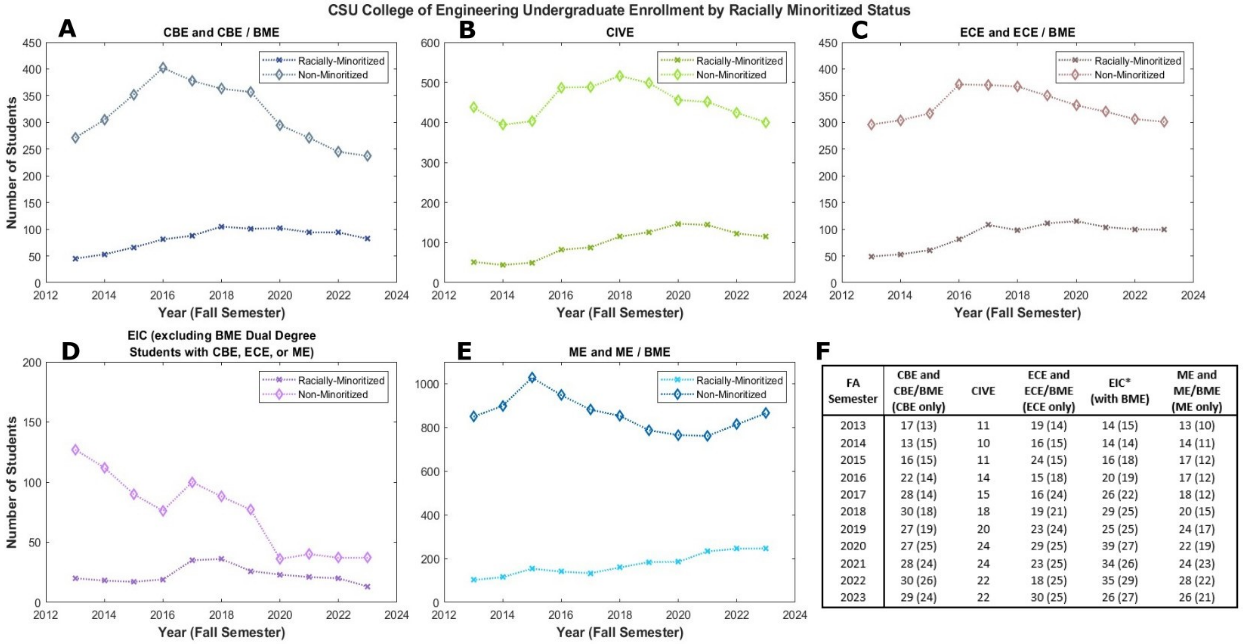


Figure 5.3: Total undergraduate enrollment in the COE at CSU from Fall 2013 to present by department, including for dual degree BME students. The * indicates that dual degree BME students with CBE, ECE, or ME have been removed from the count for EIC [196].

the racially-minoritized undergraduate population hovered between $27-30\%$ from Fall 2017 until the current semester (Fall 2023); for ME/BME, it hovered between $20-28\%$ since Fall 2018, although it wasn't above 24% until Fall 2022. When the dual majors are considered, CIVE has a considerably lower racially-minoritized undergraduate student population as compared to the other engineering departments. When considering the percentage of racially-minoritized males in each department (including BME dual degree students), the percentages in Fall 2023 are 26.7% for CBE/BME, 19.8% for CIVE, 48.6% for ECE/BME, and 11.8% for ME [196].

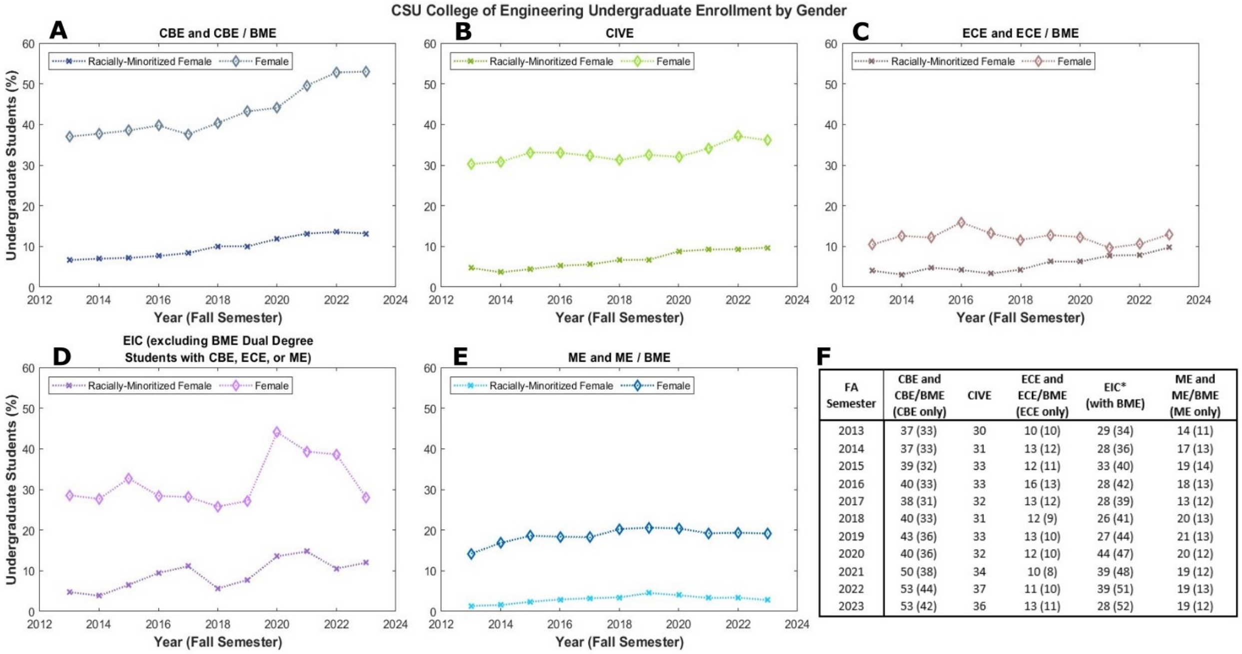
Fig. 5.5A-E show percentage of total and racially-minoritized female undergraduate students in each department from Fall 2013 to Fall 2023, and Fig. 5.5F shows the total female percentage with and without the dual degree BME students included in their respective corresponding department. From Fig. 5.5, it is clear that CBE and CIVE have made substantial gains in the overall percentage



FA Semester	CBE and CBE/BME (CBE only)	CIVE	ECE and ECE/BME (ECE only)	EIC* (with BME)	ME and ME/BME (ME only)
2013	17 (13)	11	19 (14)	14 (15)	13 (10)
2014	13 (15)	10	16 (15)	14 (14)	14 (11)
2015	16 (15)	11	24 (15)	16 (18)	17 (12)
2016	22 (14)	14	15 (18)	20 (19)	17 (12)
2017	28 (14)	15	16 (24)	26 (22)	18 (12)
2018	30 (18)	18	19 (21)	29 (25)	20 (15)
2019	27 (19)	20	23 (24)	25 (25)	24 (17)
2020	27 (25)	24	29 (25)	39 (27)	22 (19)
2021	28 (24)	24	23 (25)	34 (26)	24 (23)
2022	30 (26)	22	18 (25)	35 (29)	28 (22)
2023	29 (24)	22	30 (25)	26 (27)	26 (21)

Figure 5.4: (A)-(E): Number of racially-minoritized students in each primary department (including dual degree BME students) in the COE at CSU from 2013 to 2023. Note, the y-scale is not the same on all figures due to broad distribution in enrollment numbers. (F) The enrollment percentage of racially-minoritized students based on undergraduate enrollment for the individual department. The values in parentheses are the percentages when BME dual degree students are not included in their respective dual major programs, but instead are accounted for in Engineering Intra-College (EIC). The indicator * indicates that dual major BME students with CBE, ECE, or ME have been removed from the count for EIC. The value in parentheses include for these students in EIC [196].

of female undergraduate students overall and now have a considerably higher percentage than the ECE and ME departments, even when including the dual degree BME students with their respective corresponding departments. When comparing the percentage of female students that are also racially-minoritized (based on the overall undergraduate department population), the CBE department has the highest, with 13.2% racially-minoritized female students in Fall 2023, while the ME department has the lowest, with 2.8% racially-minoritized female students in Fall 2023. However, when the percentage of racially-minoritized female students is determined based solely on the number of *female* undergraduate students in each department (including the dual degree BME students), the ECE department has the highest percentage (75%) and ME has the lowest percentage (14.6%). The ECE department has not increased the percentage of females since Fall 2013, though, with only 13% female undergraduate students in Fall 2023.



FA Semester	CBE and CBE/BME (CBE only)	CIVE	ECE and ECE/BME (ECE only)	EIC* (with BME)	ME and ME/BME (ME only)
2013	37 (33)	30	10 (10)	29 (34)	14 (11)
2014	37 (33)	31	13 (12)	28 (36)	17 (13)
2015	39 (32)	33	12 (11)	33 (40)	19 (14)
2016	40 (33)	33	16 (13)	28 (42)	18 (13)
2017	38 (31)	32	13 (12)	28 (39)	13 (12)
2018	40 (33)	31	12 (9)	26 (41)	20 (13)
2019	43 (36)	33	13 (10)	27 (44)	21 (13)
2020	40 (36)	32	12 (10)	44 (47)	20 (12)
2021	50 (38)	34	10 (8)	39 (48)	19 (12)
2022	53 (44)	37	11 (10)	39 (51)	19 (13)
2023	53 (42)	36	13 (11)	28 (52)	19 (12)

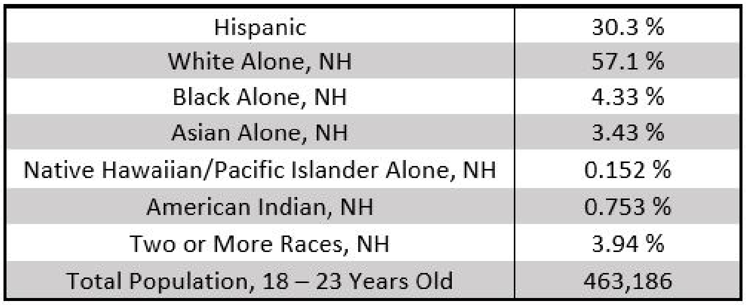
Figure 5.5: (A)-(E): Percentage of total racially-minoritized female undergraduate students in each primary department (including dual degree BME students) in the COE at CSU from 2013 to 2023. (F) The percentage of female students based on undergraduate enrollment for the individual department. The values in parentheses are the percentages when BME dual degree students are not included in their respective dual major programs, but instead are accounted for in Engineering Intra-College (EIC). The indicator * indicates that dual major BME students with CBE, ECE, or ME have been removed from the count for EIC. The value in parentheses include for these students in EIC [196].

5.3.2 Department of CBE Demographics

Beyond just comparisons to the other engineering departments, the Department of Chemical and Biological Engineering (CBE) is considered in more detail to see where they currently stand as compared to the national statistics and the State of Colorado. When looking at the CBE department individually, it has done better regarding gender balance than the national statistics show, and the percentage of females enrolled in the undergraduate program over time has generally trended upwards (see Fig. 5.6A and 5.7A), although there has been a 2.1% decrease in the percentage of female undergraduate students enrolled in the CBE program alone from Fall 2022 to Fall 2023 (from 43.6% to 41.5%). Only when the dual degree CBE/BME student population is included (Fig. 5.5A and 5.5F) is gender parity recently achieved; although the overall number of students in the CBE department is higher, the percentage of females in the department noticeably increases overall compared to CBE only, with a slightly higher percentage female than male in Fall 2022 and Fall 2023. Despite this, the CBE department has not made progress towards gender parity or adequate female faculty representation as compared to undergraduate student enrollment percentages, with only 3 females out of 17 faculty members (or 17.6%) for the 2023-2024 year [196]. The CBE department is also not doing as well faculty gender parity as the University of Colorado at Boulder (CU-Boulder), which has been hovering at or just above 30% for the last 10 years [200].

Regarding the enrollment percentages of students in racially and ethnically minoritized groups, which can be seen in Fig. 5.6, 5.4A, 5.7B-C, and 5.8, there has not been a consistent increase over time; even though there has been an increase in enrollment in some of the racially underrepresented groups, there has been a decrease in others. When the international students are removed from the statistics, as show in Fig. 5.8, the decrease in enrollment of many racially underrepresented groups is even more apparent, particularly the percentage of hispanic/latino female undergraduate students enrolled since 2018. Additionally, when compared to the racial/ethnic distribution in Colorado in 2022 for 18-23 year olds [201], the percentage of white (non-Hispanic) students is more than 10% higher than the state average in many recent years; the percentage of Hispanic/Latino students is more than 50% lower than the state average, and the percentage of black students is 40% lower than the state average.

Table 5.8: 2022 Colorado racial / ethnic demographics for 18-23 year old. [201].



Hispanic	30.3 %
White Alone, NH	57.1 %
Black Alone, NH	4.33 %
Asian Alone, NH	3.43 %
Native Hawaiian/Pacific Islander Alone, NH	0.152 %
American Indian, NH	0.753 %
Two or More Races, NH	3.94 %
Total Population, 18 – 23 Years Old	463,186

Another historically underrepresented and marginalized group is first generation students, shown in Fig. 5.9, with enrollment generally trending upwards to $22 - 23\%$ before the Fall 2020

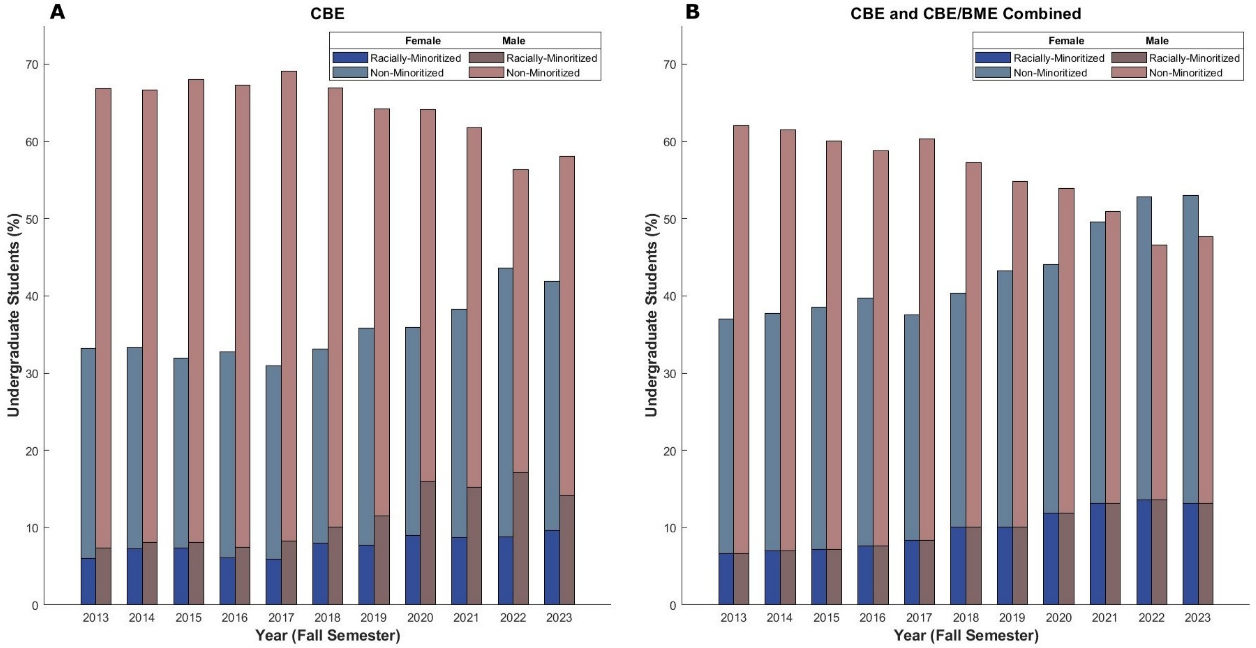


Figure 5.6: Percentage distribution by gender and racially-minoritized status for the (A) CBE program and (B) dual degree CBE/BME program from the Fall 2013 semester to present [196, 198].

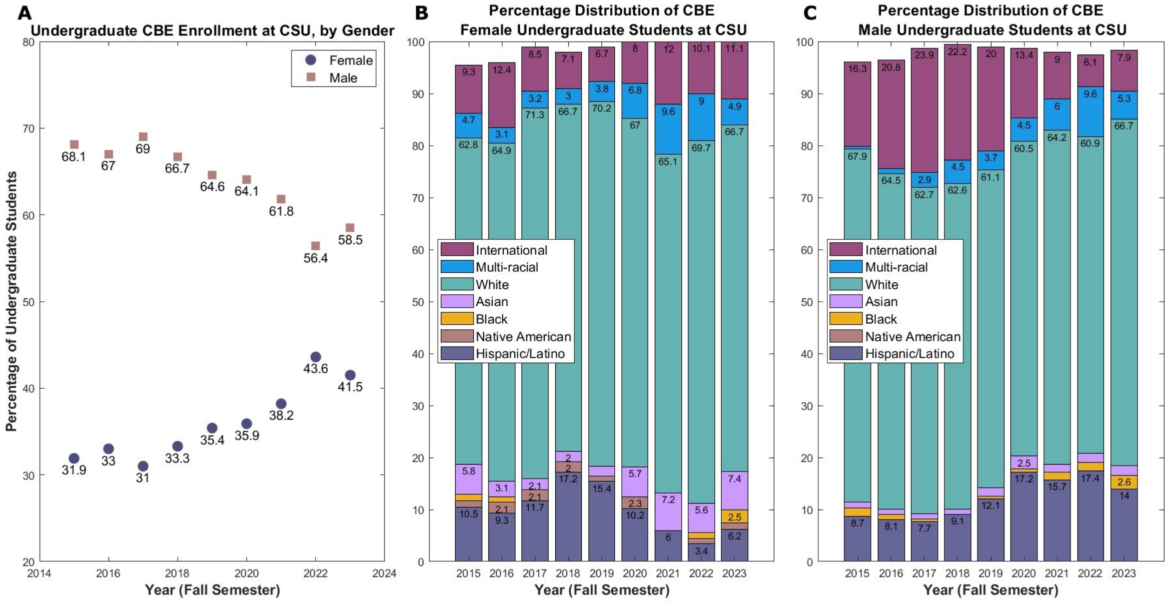


Figure 5.7: Percentage distribution of CBE undergraduate enrollment demographics at Colorado State University by gender and ethnicity from Fall 2015 to present. For (B) and (C), numeric values less than 2% are not explicitly displayed on bars, and "no response" regarding ethnicity are excluded from figures. In Fall 2018, there was one male of Hawaiian/Pacific Islander ethnicity that is not shown in (C) [196, 198].

semester (with only a slight dip in Fall 2019 compared to the previous year); however, enrollment of first generation students appears to have been dramatically impacted by the Covid-19 pandemic,

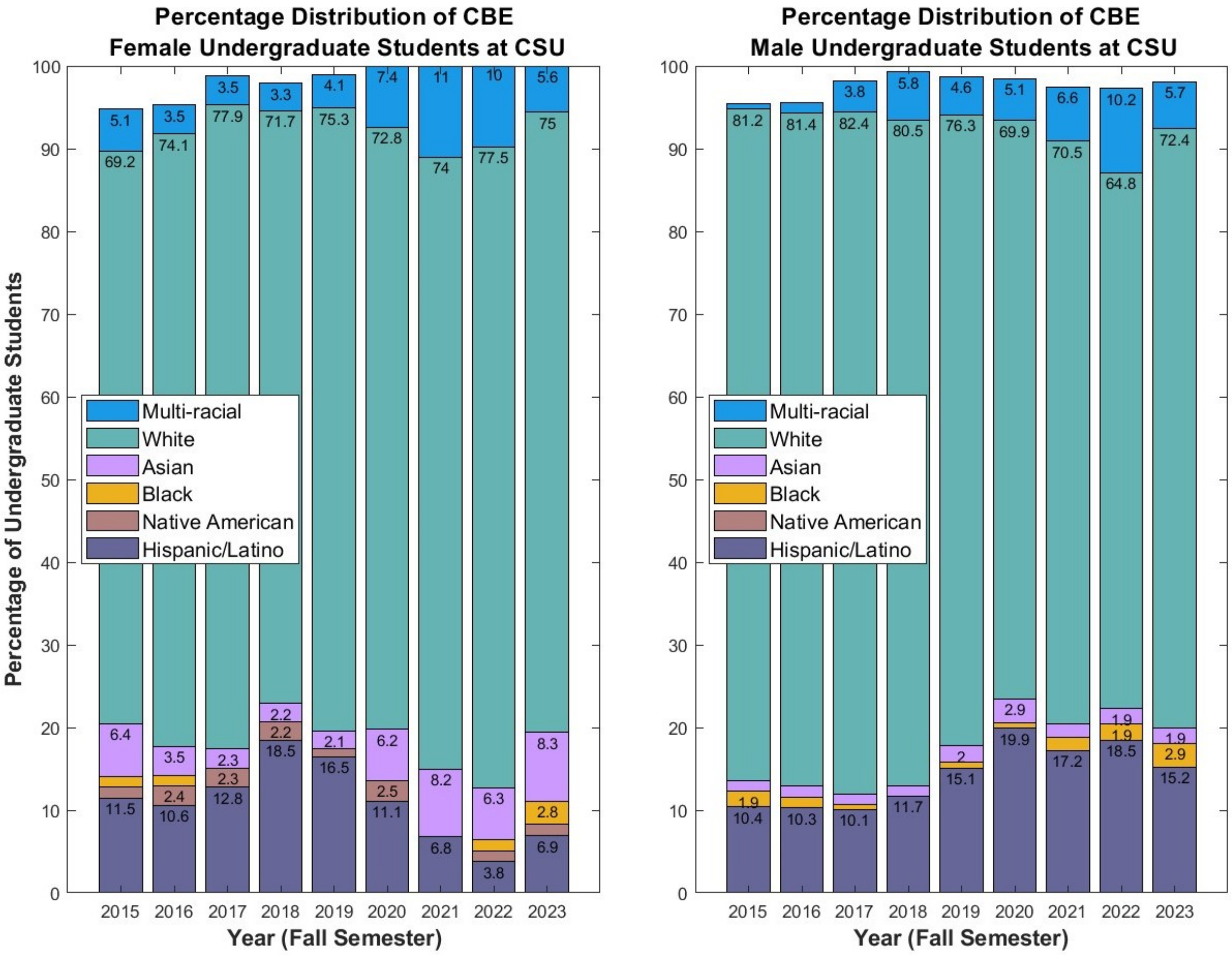


Figure 5.8: Percentage distribution of CBE undergraduate enrollment demographics (without international students included) at Colorado State University by gender and ethnicity from Fall 2015 to present. Numeric values less than 1.8% are not explicitly displayed on bars, and "no response" regarding ethnicity are excluded from figures. In Fall 2018, there was one male of Hawaiian/Pacific Islander ethnicity that is not shown in (C) [196, 198].

decreasing to 14.7% by Fall 2022, a level not seen since 2011. At present (Fall 2023), this has rebounded to 18.5%, but whether this trend continues remains to be seen. Unfortunately, 50% of the first generation students in the undergraduate CBE program are seniors, so it is possible the number of first generation students in the program will decrease considerably by Fall 2024 unless there is a significant increase in freshman enrollment of first generation students.

Fig. 5.10 shows the undergraduate persistence rates by cohort for CBE only students as well as dual degree BME/CBE students from Fall 2015 through Fall 2022. When looking at Fig. 5.10A compared to Fig. 5.10B and C, persistence rates for the 2018 and 2019 cohorts appear to have been

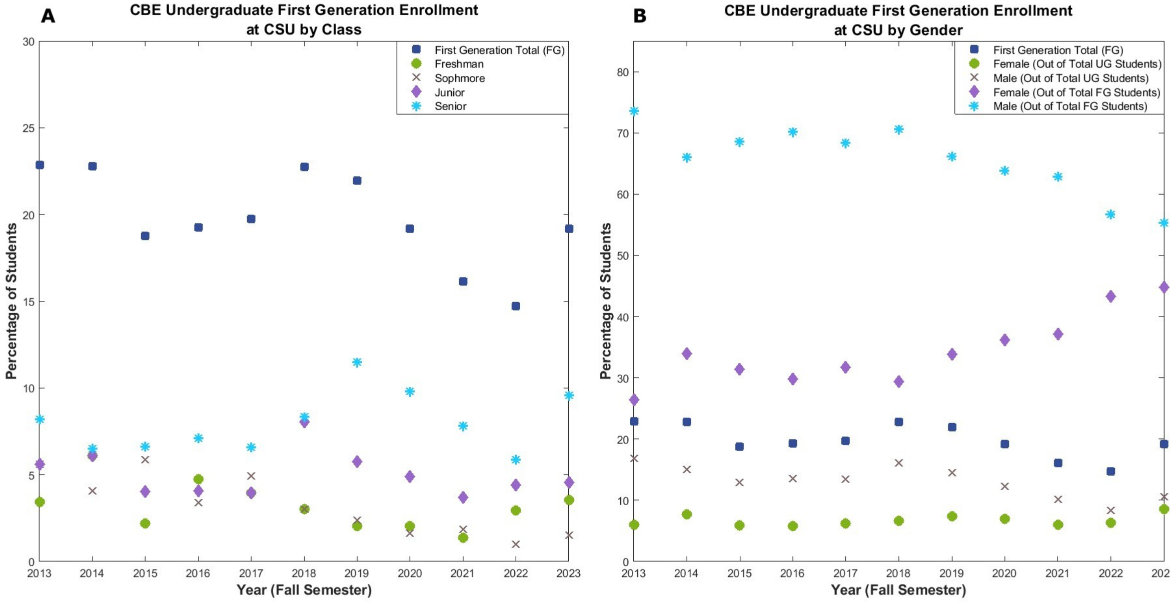


Figure 5.9: Percentage distribution of first generation undergraduate students in CBE at CSU by (A) class and (B) gender from Fall 2013 to present. Note: Students classified as post Bachelors are excluded [196].

negatively impacted by the Covid-19 pandemic; however, the primary negative impact of Covid-19 may have been on enrollment and not persistence, as demonstrated by the numbers for CBE only students and dual degree CBE/BME students in the 5.10C legend. Although persistence appears to be lower for the CBE only students in the Fall 2022 cohort, it is similar to historic data.

An issue did become apparent upon analysis, however, with how this data is reported. Although data regarding persistence within the University, College, or Department are available, since the dual degree students are not technically in their corresponding department, students switching from CBE to the dual degree CBE/BME or vice versa are not accounted for within the persistence data, which may make persistence numbers appear lower than they actually are. This can be seen when comparing persistence data to the *Major Migration Between Entering Cohort and 3rd Fall* data, which unfortunately is only available for Cohort Fall Terms of Fall 2018, Fall 2019, and Fall 2020.

Looking at the *Major Migration Between Entering Cohort and 3rd Fall* data for the Fall 2020 cohort, shown in shown in Table 5.9, 69.6% of CBE only students and 69.4% dual degree BME/CBE students were still in the same program. However, 3 of the 4 students that left the BME/CBE program for another program in the same college switched to CBE only. When adjusting for this, the department more accurately had a retention rate for this overall cohort of 74.6% during this time frame. Of the students that left the CBE program, roughly two thirds transferred to programs outside of the College of Engineering and the other third of students left the university overall. Additionally, the total Fall 2020 cohort gained 12 additional students from different programs during this time, so it is currently 89.8% of its original size. The numbers were not as favorable for the Fall 2018 and Fall 2019 (CBE and CBE/BME combined) cohorts, however, with adjusted retention rates of 47.0% and 56.5%, respectively. Although these cohorts were large to begin with, even with students transferring into the program, they were only 65.0% and 75.0% of their original size, respectively, by the 3rd fall semester.

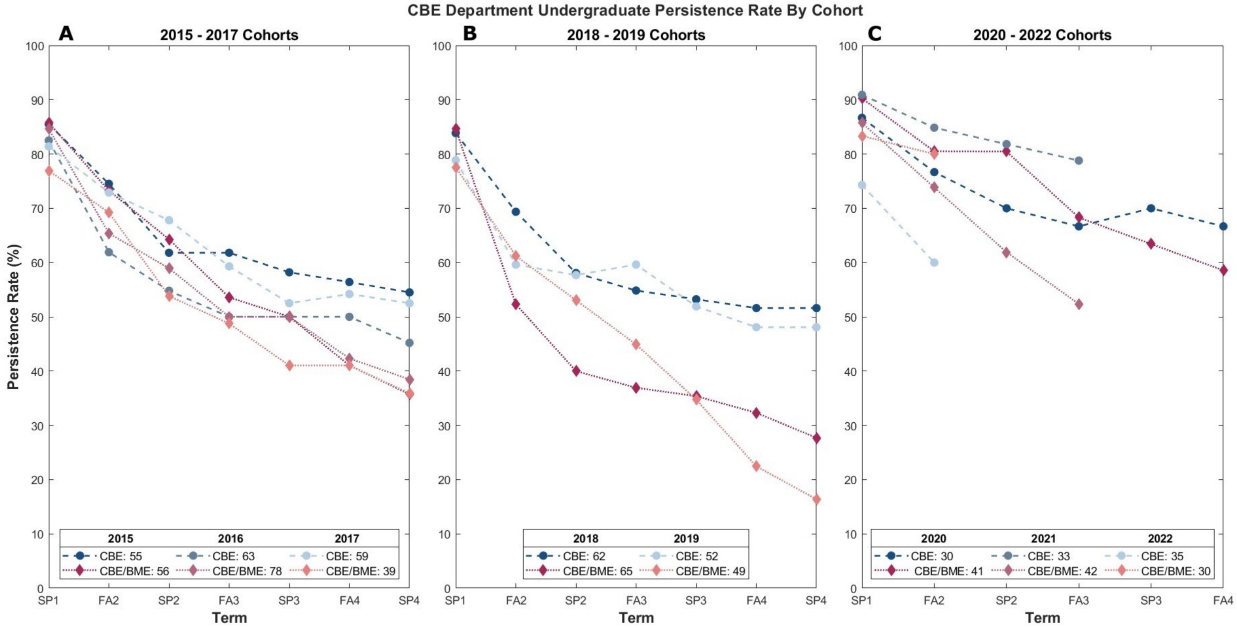
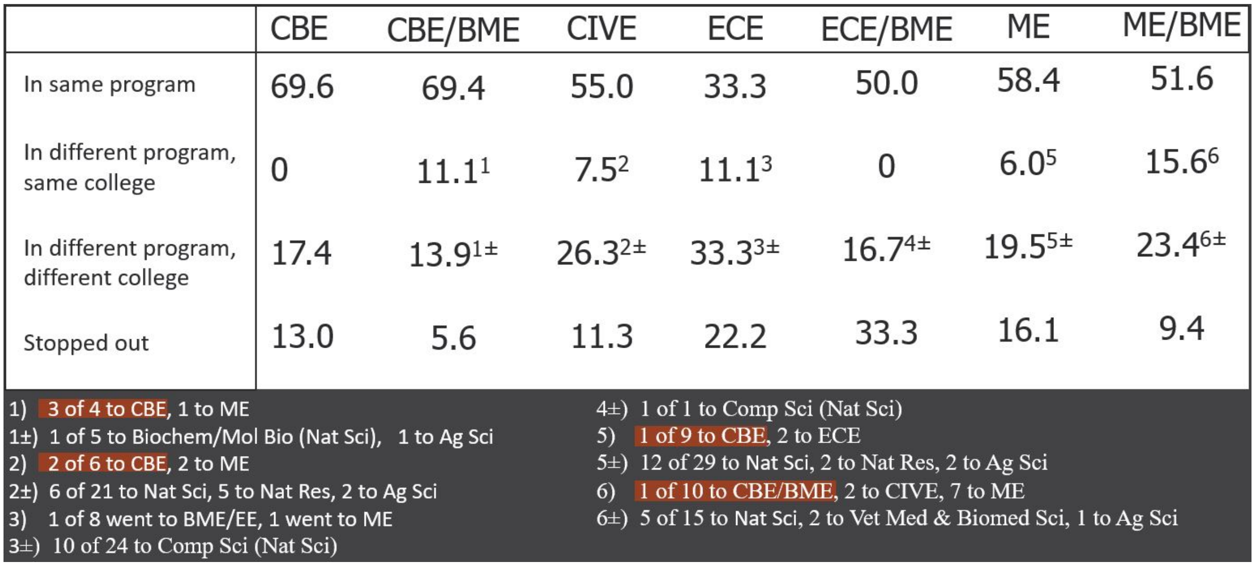


Figure 5.10: CBE and CBE/BME Department Undergraduate Persistence Rate from Fall 2015 to Present, by Cohort. [196].

In addition, as seen in Table 5.1 and Fig. 5.3, the overall number of students enrolled in the CBE department (both single major and dual degree with BME) is lower than overall enrollment in other engineering departments since Fall 2019, possibly as a result of the Covid-19 pandemic, even without taking into consideration any of their dual degree undergraduate students, with the number of students in CBE decreasing to levels not seen since before 2009. Given the enrollment chal-

Table 5.9: The major migration percentage between the Fall 2020 entering cohort and the third fall, by COE Department at CSU. For the CBE, ECE, and ME departments, the percentage of students in the department is shown separately than the dual degree BME students. [196].



	CBE	CBE/BME	CIVE	ECE	ECE/BME	ME	ME/BME
In same program	69.6	69.4	55.0	33.3	50.0	58.4	51.6
In different program, same college	0	11.1[1]	7.5[2]	11.1[3]	0	6.0[5]	15.6[6]
In different program, different college	17.4	13.9[1±]	26.3[2±]	33.3[3±]	16.7[4±]	19.5[5±]	23.4[6±]
Stopped out	13.0	5.6	11.3	22.2	33.3	16.1	9.4

1) 3 of 4 to CBE, 1 to ME
1±) 1 of 5 to Biochem/Mol Bio (Nat Sci), 1 to Ag Sci
2) 2 of 6 to CBE, 2 to ME
2±) 6 of 21 to Nat Sci, 5 to Nat Res, 2 to Ag Sci
3) 1 of 8 went to BME/EE, 1 went to ME
3±) 10 of 24 to Comp Sci (Nat Sci)
4±) 1 of 1 to Comp Sci (Nat Sci)
5) 1 of 9 to CBE, 2 to ECE
5±) 12 of 29 to Nat Sci, 2 to Nat Res, 2 to Ag Sci
6) 1 of 10 to CBE/BME, 2 to CIVE, 7 to ME
6±) 5 of 15 to Nat Sci, 2 to Vet Med & Biomed Sci, 1 to Ag Sci

lenges and current representation levels of underrepresented and historically excluded populations at both CSU and in the CBE department, the CBE department needs to increase the enrollment of these students as well as improve overall enrollment in the department. To do this, it is crucial that the department attract students from diverse backgrounds, but also make changes within the program and adapt the program to be more inclusive overall and support these students, including through an increase in the presence of DEI initiatives in the undergraduate CBE curriculum, which will likely positively contribute to an increase in retention throughout the undergraduate population.

5.4 Previous Work and Lessons Learned

As part of a multi-year NSF-supported initiative, a variety of approaches to incorporate multiple activities and assignments in the first- and second-year CBE courses have previously been implemented in order to help students improve teamwork and learn about implicit biases, with an aim of assisting students in fostering more inclusive professional identities and recognizing the value of diversity and inclusion in engineering. These courses, along with the other required courses within the CBE department, are shown in Fig. 5.11. Even though other courses are also required for all CBE students, those courses are not offered within the CBE department and are, therefore, not taught by CBE faculty and the curriculum and pedagogy are not within the department's control. This work was presented at the 2021 Collaborative Network for Engineering and Computing Diversity (CoNECD) Conference [168]. Our long-term goal in this endeavor is to incorporate topics related to DEI throughout all four years of the undergraduate CBE curriculum (shown in Fig. 5.11) in the future. The approach utilized in each individual course has been (and will continue to be) dependent upon both the structure and the content of the course. In addition, we have taken a variety of approaches in different courses in an effort to improve student participation by decreasing repetition across courses and redundancy in assignments and activities.

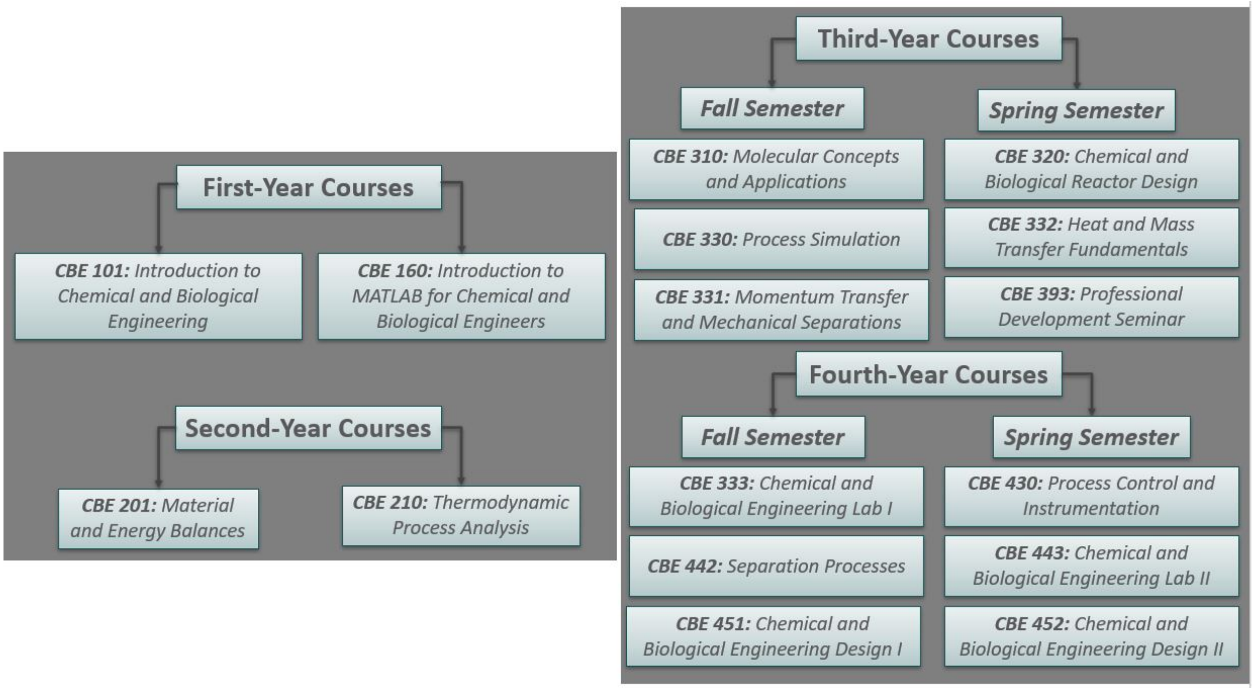


Figure 5.11: These required courses for the CBE Program are offered and taught by the CBE department at CSU. Even though other courses are also required for all CBE students, those courses are not offered within the CBE department and are, therefore, not taught by CBE faculty and the curriculum and pedagogy are not within the department's control. [168].

When considering possibilities for the undergraduate curriculum, it is may be helpful to consider DEI initiatives from multiple different angles. There are specific types of knowledge and skills that can help individuals recognize cultural barriers to inclusion and work more effectively with difference. For example, every person has implicit biases which can impact their thoughts, actions, behaviors, and decisions towards others unconsciously in spite of the person's conscious external values and beliefs [202]. It is, therefore, vital to educate students about implicit biases and their ability to unconsciously impact one's thoughts and actions. Additionally, the ability to work well in teams and effectively collaborate with others, which require interpersonal skills that can be developed and/or improved, are essential for an engineer in both academia and a professional setting, regardless of industry [189, 203–206]. Building a strong interpersonal skill set entails recognizing the value every individual and their unique perspective contributes to the team, having an understanding of personal strengths and weaknesses, clearly and effectively communicating ideas, actively listening to the ideas of others, and exhibiting sensitivity regarding differences. More broadly, an inclusive mindset and an appreciation for the value of diversity within the group are critical factors that contribute to a team being successful and achieving its goals. While the need for engineering students to be able to work in teams has long been recognized, engineering education has not always recognized the need to explicitly teach about teamwork and sought to develop teamwork as a skill. Incorporating diversity and inclusion topics so that students learn the crucial interpersonal skills necessary to foster a supportive and respectful team environment, however, can be a challenge in an academic engineering environment in which nearly all of the course content is highly technical. Simply placing students in teams for project work and laboratories is insufficient without also providing students with the knowledge and tools required to enable them to gain the interpersonal skills that are needed for more successful interpersonal interactions, including educating them on implicit biases [192, 207]. Incorporating these topics throughout the undergraduate engineering education alongside their technical core engineering content can not only provide students with the necessary tools to achieve more effective academic teams, but also enable them to cultivate more inclusive professional identities and gain additional skills that will be beneficial long-term in their careers. Although effective teamwork and implicit bias are just two components of the broader concepts of diversity and inclusion in engineering, the ease with which they can be incorporated into both non-technical and technical course content make them a good starting point for introducing these concepts in engineering curricula.

5.4.1 First-Year Courses

CBE 101: Introduction to Chemical and Biological Engineering

In the first-year course, Introduction to Chemical and Biological Engineering (CBE 101), we included multiple diversity related topics because the structure of the course more easily allows for the addition of non-technical course content. An interactive theatre sketch of a dysfunctional team interaction has been performed, with trained facilitators guiding the activity [189–191]. In the interactive theatre sketch activity, students watch the full interaction between dysfunctional team members. The class discusses the interaction and negative behaviors and treatment between members. The group then starts the sketch over again, but this time the students have an opportunity to stop the interaction at a specific point and join in to contribute to changing the direction, language, or intensity of the interaction to improve respect and inclusivity. This was included in the CBE 101 course starting in 2017 and has continued since then, although it is only incorporated in the fall semester of each academic year. In addition, I included it in CBE 101A course when I taught it in Fall 2022; at this time, the theatre troupe was funded by the Walter Scott, Jr. College of Engineering, thanks to the efforts of Melissa Burt. However, a significant challenge to routinely incorporating this into the course is funding, since hiring the theatre troupe is quite expensive.

To introduce the concept of implicit bias, we incorporated an activity in the lab sections of the CBE 101 course, which enabled us to implement the activity for smaller groups of students. In this activity, the students watched a video introducing implicit bias, after which they took multiple Harvard Implicit Association Tests (IATs), participated in a group discussion, and answered reflection questions regarding their test results [191]. The implicit bias activity was implemented in the course in many semesters between Fall 2017 and Spring 2020. It was discontinued, however, in following years for a couple of reasons. When we incorporated the activity initially, we discussed how the results are not necessarily predictive and conclusive, but rather just a tool to use over time to get students thinking about implicit biases. However, as time has progressed, a growing body of research regarding the complex nature of implicit versus explicit behaviors and the associated contexts as well as what the IAT is actually measuring, and shortcomings and flaws associated with the IAT [208–211]. At the same time as changes were being considered to the activity, the Covid-19 pandemic affected the format of instruction in courses, with hybrid and/or online components, so retiring the activity was the best approach. Regular surveys were also incorporated into the course to assess these various activities.

An additional activity that was incorporated into the CBE 101 course in Spring 2022, Fall 2022, and Spring 2023 was a 50-minutes Science Communication Workshop [212–214], presented by Nicole Kelp. The students take a survey before the workshop and another roughly a month afterward. The workshop is an interactive experience in which the students have to communicate regarding different scientific topics to individuals with different backgrounds and careers. This workshop not only encouraged students to consider how to communicate scientific topics, but also encouraged engagement and positive socialization between the students in the classroom. Some results from the pre- and post-surveys done for this workshop for courses in multiple STEM departments and programs at CSU are discussed in Alderfer, *et al.* (2023) [212].

Related to DEI in the pedagogy and course structure itself, numerous changes were implemented from Spring 2022 through Spring 2023. The course syllabus was restructured to include the following: an Inclusive Classroom statement, an extended Universal Design for Learning / Accommodation of Needs, an Undocumented Student Support statement, Religious Observations statement, and Basic Needs statement. The homework policy was adapted after Spring 2022 to include the acceptance of late HW with a daily penalty, with further flexibility if students reached out to explain their circumstances. I also substantially increased the number of learning assistant help hours in addition to my help hours, as well as took a poll in order to have help hours that were varied enough to accommodate as many students as possible. Overall, many students seemed to respond well to the changes. For example, some anonymous student comments from Spring 2023 include:

Student 1*: The professor was really understanding when it came to extensions and getting help*

Student 2*: The concepts were fairly hard to grasp since it is nothing like what iv'e seen before. There is an abundance of resources though to help understand the concepts. Ms. Weber gives lecture notes in person and in online videos. We have access to the book which is helpful and whenever there is confusion Ms. Weber explains well. We also have countless help hours.*

Student 3*: The classroom environment was very inclusive and helpful to my learning. The instructor conveyed the concepts clearly, but could have added more examples problems to go more in depth with the concepts...*

Student 4*: Ms. Weber seems to be very knowledgeable with these subjects and explains them well. She also was very understanding of our schedules most of the time and would push back homework if need be. I also liked when she would give examples of when bad engineering out in the work force created a problem,...*

CBE 160: Introduction to MATLAB for Chemical and Biological Engineers

In the first-year computing course, Introduction to MATLAB for Chemical and Biological Engineers (CBE 160), many changes have been made to incorporate DEI into the class, both in terms of the curriculum and pedagogy. Related to DEI in the pedagogy and course structure itself, numerous changes have been implemented. The course syllabus has been restructured and now includes the following: an Inclusive Classroom statement, an extended Universal Design for Learning / Accommodation of Needs, an Undocumented Student Support statement, Religious Observations statement, and Basic Needs statement. In addition, Canvas for the course has been restructured to have easier-to-navigate homepage and weekly modules with objectives, reminders of reading and assignments, course files, and posting of class session videos (edited to remove any downtime during class sessions). Further, at the beginning of the semester, an instruction guide is provided to students via email and on Canvas (as an announcement and preparation module that explains, with numerous links included, how to navigate the Engineering Technology Services (ETS) website to create an engineering account, set up personal computers, access their university drive, download software, use the VPN, etc. A QR code for tutoring resources through the COE and a MATLAB HW assignment template are also provided for students. Additionally, an Introductions Discussion Board is started on Canvas before each semester begins for students to engage introduce themselves to the class, specify a preferred name and identify preferred pronouns (if desired), and list at least one random fact about themselves, and they are encouraged to respond to at least one other student's post. The purpose of this is for students to begin the process of engaging with and getting to know one another before they meet each other in person. Each week, partially completed weekly notes (in MATLAB) are provided to students so they can participate and stay engaged in class while practicing coding and the topics we are covering, but also not be hindered by an inability to type quickly. Typically, half to three quarters of each class session is dedicated to course notes, but the rest of class time is dedicated to working together, helping one another, and receiving help from me on "in-class problems" that are some of challenging questions on the HW assignment. This helps with students understanding of the material as well as with building relationships with their classmates and having learning be more of a collaborative process. In addition to these changes, changes have been made on the grading to:

1. drop two HW assignment scores, no questions asked;
2. drop one weekly (in-person) quiz, no questions asked;
3. provide some quiz accommodations to allow a student to take a weekly quiz remotely if they have a situation that comes up regarding an illness or other circumstance (with some requirements, particularly regarding communicating with me in advance of missing class and that the policy has not been taken advantage of);
4. and provide quiz extra credit assignments for more than 3/4 of the weekly quizzes that are designed to for any student to take (whether they missed the quiz, received a 0, or received a 100%) to increase their understanding of the material (that overall are worth nearly an entire additional quiz).

The purpose of these numerous class policy changes and course structure changes is to account for circumstances beyond a student's control, to improve equity by recognizing that different students have different life circumstances or obstacles outside of their academics that may interfere with their classes, and to provide some understanding regarding the different ways in which different students learn. The student feedback regarding these changes was quite positive. For example, from Fall 2022, some anonymous student comments include:

Student 1*: Ms. Weber really went in depth and made sure we all were following the information. The classroom was very open and comfortable making it easy to ask questions.*

Student 2*: Some good-to-excellent aspects of the course instruction was the in class notes provided and level of care during class and office hours. The in class notes are detailed and interactive. This helps me stay engaged throughout class and also helps me learn better. The professor also provided a lot of office hour opportunities which were helpful because it offered a chance for one on one help on homework and collaboration with other students.*

Student 3*: Some good to excellent aspects of the course instruction were the in class notes because I found them to be engaging as well as very helpful when referring back to in order to complete the homework. The class is split in a way that usually works very well with most of it devoted to notes with some time at the end to work through a more challenging problem with our peers.*

Student 4*: Ms. Weber used the class periods very effectively in getting the course content across to the students. This also created a good class environment where everyone is listening to the content and asking questions when one comes up.*

From Spring 2023, some anonymous student comments include:

Student 1*: The way Ms. Weber set up the in-class notes as skeleton notes in which we had to fill out together with her during the lecture was very impactful. It kept me engaged in the very long lecture while allowing me to actively learn and recall material.*

Student 2*: I like how the instructor actually did the code alongside us in the in class notes and you could visibly see where and why you were making errors if you made them.*

Student 3*: The instructor did well on very clearly showing us what concepts are, especially by being very loud, clear, and coherant. She was very open with students asking questions/created a safe environment for students to be vulnerable with this difficult course*

Student 4: *The instructor was really great and conveyed concepts clearly, while working with students so they can do their best to succeed.*

Student 5: *The classroom environment was really good. Classes were also made to make the content clear to understand.*

As related to DEI in the curriculum, the students perform multiple assignments in which they apply the programming skills learned in the course to analyze the gender-pay gap as a result of different starting pay rates, annual raises, and promotion periods. This assignment was originally based on one developed as part of the NSF-initiative at West Virginia University. When this assignment is introduced, a discussion takes place regarding the intersectionality of pay disparities based on gender with pay disparities based on race/ethnicity as well as other factors (such as LGBTQ+ status). In this discussion, it is specified that although these other factors may also come into play in overall pay (and specifically in salaries), many of these are either more complex to analyze, more difficult to find reliable data, or both. The homework problems related to this are spread across multiple assignments and are still incorporated in the course, as of Fall 2023. The students are also asked to answer reflection questions regarding their results. The reflection questions were modified in Fall 2023; prior to this, however, the reflection questions and some examples of students responses are:

1. Was there anything surprising in the results? Explain why or why not?
2. Do you expect what you learned in this assignment will impact how you interact with other engineering students or engineers in your profession, especially those who are different from you? If so, how?
3. Did what you learned in this assignment change your views on the roles and responsibilities of engineers? If so, how?
4. After completing this assignment, do you feel like you have a better understanding of coding loops in MATLAB and/or applications of MATLAB? Please explain.

Student 1: *What I learned in this assignment will definitely impact how I react with engineers especially when it comes to future bosses and fighting for equal pay and knowing my worth as an* *engineer.The assignment did change the roles and views of engineers. I think they have the role now to stop the pay gap and encourage a diverse and healthy working enviornment. I feel like I have learned that Matlab can be used for realworld applications instead of just math problems, and real world problems can have real implications from the data produced by matlab.*

Student 2*: I didn't realize how much a women's salary could differ from a man's salary. It makes me sad that women can work just as hard or even harder than men, and not get rewarded for it. I might be biased since I am a women but if I was doing the same work as a man and he got paid better than I did, I would be very upset. Yes it did. Since I didn't realize that pay gap was as serious of an issue as this code proves it to be, I find it more annoying to be in engineering as a woman because I will most likely get paid worse than my colleagues. Engineers still have the same roles and responsibilities but I believe that we should be working to fix this problem and pay people when people are due and deserve it.*

Possible improvements for this assignment include the addition of pay disparities based on race/ethnicity or adding reflection questions after each portion of the assignment. Possible improvements for this course include adding one or two mini projects (or making a couple of the homework assignments group assignments instead of individual) in the semester as well as adding in reflection questions regarding working in these groups. Although students currently do group evaluations at the end of the final project, most students add reflections into these and it may be beneficial to formally add them in as a separate survey or question set as well.

5.4.2 Second-Year Courses

In the second-year courses, Material and Energy Balances (CBE 201) as well as Thermodynamics (CBE 210), incorporating DEI topics is more challenging as a result of the highly technical course content. The frequent group projects in these courses allowed for incorporation of diversity related topics with a series of reflection questions after each project that address individual behavior within the group, interactions with team members, demonstration of respect for group members' contributions, and overall inclusion of group members. Additionally, this was an easily implemented strategy for enabling students to learn effective teamwork skills while not infringing on the technical course content. Reflecting on their experiences interacting with their classmates and working in groups can provide students' insight into their individual strengths and weaknesses, help them in communicating their ideas more effectively, and can provide improved understanding into what skills, actions, and strategies are useful for creating a more open, supportive, and respectful team environment. These questions were designed to promote a growth mindset in how the students view their individual group contributions, as well as interactions within the group. It is believed that the reflection questions are still incorporated in CBE 201 to-date; in CBE 210, they continued until at least Spring 2021, and whether they are still incorporated is unknown.

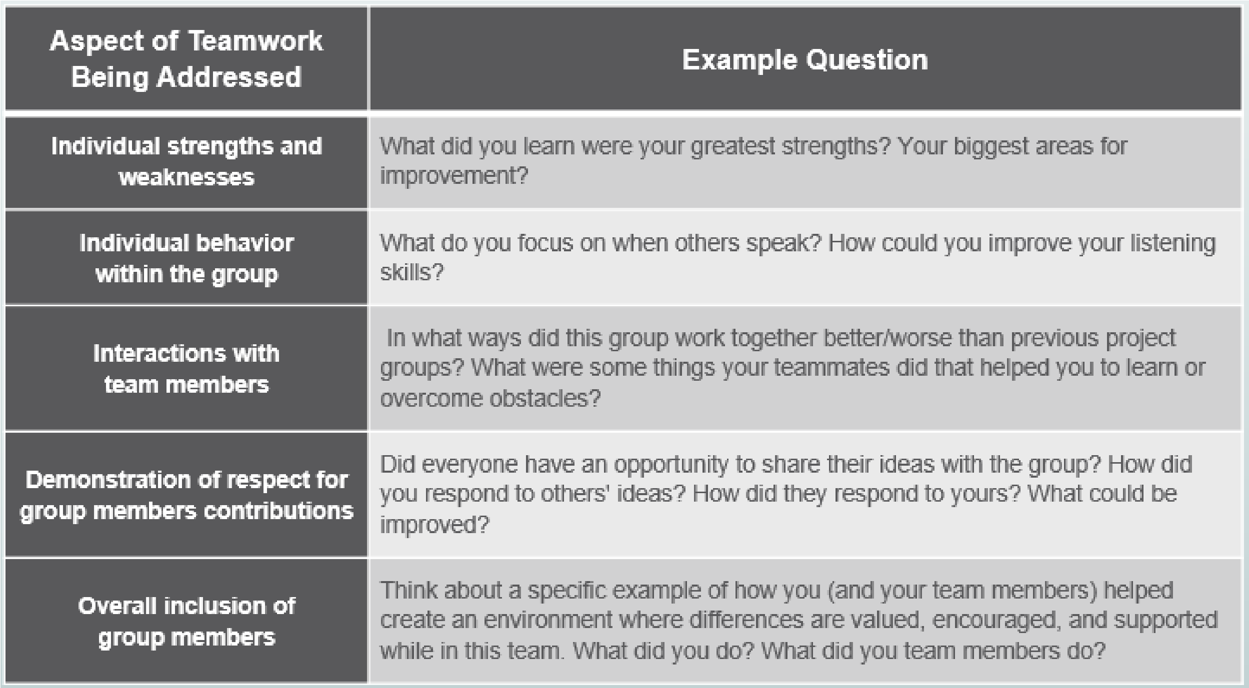
CBE 201: Material and Energy Balances (CBE 201)

For CBE 201, four group projects are completed throughout the semester. For each of these projects, students are in a new group. This provided us with an opportunity to reflect on their interactions after each one to reflect on the experience, possibly identify elements that went well or could be improved, and compare with the previous project teams. The students also completed a final set of questions reflecting on the overall team experiences throughout the semester. Each set of reflection questions provided a small amount of extra credit in the course. In this course, the reflection questions were designed to address multiple aspects of individual work within a team as well as teamwork overall. These aspects of working in teams as well as a related reflection question example are shown in Table 5.10. The reflection questions were selected and grouped strategically for each group project to address all of the following: Individual behavior (self reflection), Teamwork (group reflection), and Inclusivity (inclusion, appreciate for differences, and respect for group members).

In Fall 2019, 82 students completed all four sets of project reflection questions as well as the final set of reflections; in Fall 2020, 63 students completed all five total sets of reflection questions. An examples of one of the reflection questions asked after completing the first group project and examples of student responses are:

P1 Q2: What is the most valuable skill you learned from your teammates / project group that you

Table 5.10: Examples of the CBE 201 group project reflection question, as related to the aspect of working in teams they are intended to address [168].



Aspect of Teamwork Being Addressed	Example Question
Individual strengths and weaknesses	What did you learn were your greatest strengths? Your biggest areas for improvement?
Individual behavior within the group	What do you focus on when others speak? How could you improve your listening skills?
Interactions with team members	In what ways did this group work together better/worse than previous project groups? What were some things your teammates did that helped you to learn or overcome obstacles?
Demonstration of respect for group members contributions	Did everyone have an opportunity to share their ideas with the group? How did you respond to others' ideas? How did they respond to yours? What could be improved?
Overall inclusion of group members	Think about a specific example of how you (and your team members) helped create an environment where differences are valued, encouraged, and supported while in this team. What did you do? What did you team members do?

will incorporate in future teamwork?

A1: *The most valuable skill I learned from my teammates was that it is very important to meet in person and communicating about your thoughts and questions on the project is vital. I also learned that it is ok to ask your teammates questions if are confused or don't know how to approach something. This way your team can help you and they can feel open and comfortable to ask you questions as well so you can help them too.*

A2: *I learned that patience is necessary when working with a group (especially a group of strangers), such as when extra explanation is needed.*

A3: *There are a lot of people who are better are certain subjects than me, so it's very valuable to listen to everyone's ideas.*

Each subsequent question set built upon the previous set, asking students to compare previous interactions with their most recent interactions as well as to discuss what changes they can make to improve future interactions. The subsequent question sets were also designed for students to recognize and reflect on both the positive and negative aspects of their group interactions. An example of two questions asked for Project 2 and examples of student responses are:

P2 Q2: What did you learn were your greatest strengths? Your biggest areas for improvement?

A1: *My greatest strength was actually solving the material balance problems. I need to improve my ability to communicate my work to my other teammates so that they could understand it.*

A2: *I learned that my greatest strengths were my communication, and my biggest area of improvement would probably be trying to get on the same level of understanding of everyone else.*

A3: *I think that my biggest strengths was being able to communicate my points across and properly suggest how to tackle a problem without upsetting anyone. In that the group was able to take in all ideas and select the best one possible with all voices being heard equally. I believe I need to improve on knowing the material better so that I can contribute accurate and correct information to my group as some of the ideas I put out may have not have been the best until I figured out the material that I was doing was wrong and then I had to correct it. Such as the idea that a stream with only on variable does no count as specified.*

P2 Q3: How do you ensure people know that you have taken account of their views?

A: *I acknowledge their ideas when they present them and try to make compromises out of all group members ideas. If you ask people questions about their ideas, then you also gain more knowledge about their views and they see you are validating them.*

The final set of reflection questions asked students to look back on all of their group experiences throughout the entire semester. Although most of the reflection questions are designed strictly for the benefit of the students, responses to the final question regarding topics that may benefit from formal training provide some actionable suggestions that the department could implement relatively easily, such as a short seminar on group work, communication, etc. All of these responses and as well as some identified themes are discussed in Appendix B. The final question, which relates directly to changes that could be made by the department, and some examples of student responses are:

Final Reflections Question: Experience is an important teacher; however, sometimes formal instruction or training on particular topics can be helpful. Thinking about your team experiences this semester, are there topics or types of interactions where you would benefit from formal training? In other words, if the CBE department decided to incorporate some formal team training skills into this course or later courses. are there particular topics you think that training should cover?

A1: *I think some of the people in this major could benefit from trainings on how to be respectful to other group members, especially ones that come from diverse backgrounds. One particular annoyance is the tendency for males to interrupt females while they are talking (i have experienced this many times) and to generally ignore what females are saying. For example, I, as a female will suggest one idea (that will be generally disregarded), but when that same idea is suggested by a male in the group, everyone will agree to do it.*

A2: *I think that training should cover how to respectfully get in contact with people you have never met before. It would also be good to address group etiquette and expectations.*

A3: *There are some basic skills that I am not sure would even be considered in this training but that some fellow students still lack- mostly communication. Things like reaching out to group members early on, responding to these messages, and retaining respectful language even while frustrated. But I believe one important topic that the training should actually cover is how to divide work among a team efficiently and respectfully. In the real world, we are going to be working with people of different skill levels. We need to know how to assign the tasks of a project to these people in a way that makes the work most efficient but does not make anyone feel disrespected or undervalued.*

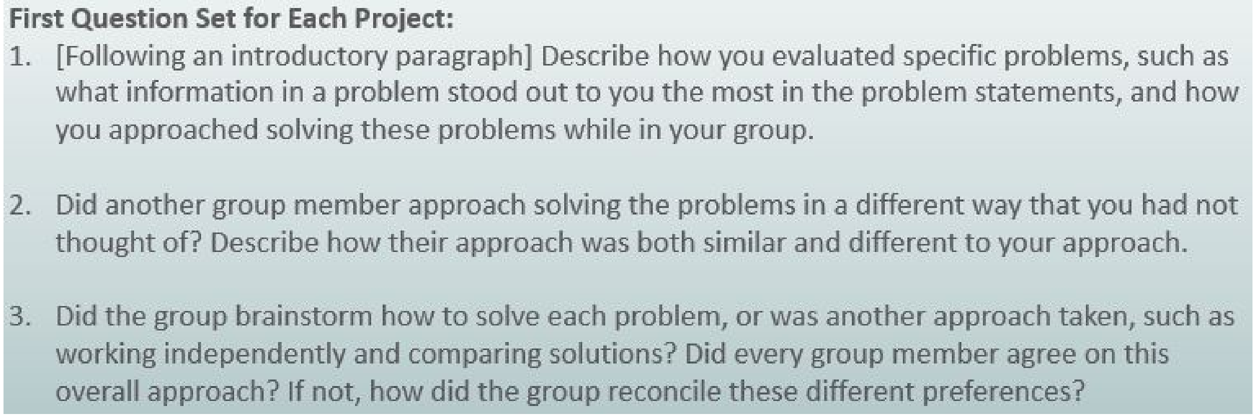
A4: *The training should cover how to properly communicate with each other. This could touch on how to communicate with your group mate's ideas in order to properly acknowledge everyone and make everyone's ideals feel validated.*

A5: *I think that a small seminar based on teamwork could be helpful. In this seminar, students should be taught about organizing meetings, distributing workloads, understanding skillsets, and working as a team. This would likely help people with their groups and group organization.*

A6: *It might be helpful to take the material and energy balances that we learn in class and apply them into real life situations. I feel like it would be a great way to help figure out why we set the balances up the way we do...*

CBE 210: Material and Energy Balances (CBE 210)

In CBE 210, the students completed group problem solving sessions each week. They rotated groups and worked with three different groups throughout the semester. The reflection questions after working with each individual group are different from those in Material and Energy Balances. These questions are more focused on developing an understanding and appreciation in students of the unique perspectives and experiences each member has, how these perspectives affect how each member approaches a problem, and how the diversity within the group can help the group to more effectively achieve its goals For example, the first question set for each project is shown in Fig. 5.12.



First Question Set for Each Project:

1. [Following an introductory paragraph] Describe how you evaluated specific problems, such as what information in a problem stood out to you the most in the problem statements, and how you approached solving these problems while in your group.

2. Did another group member approach solving the problems in a different way that you had not thought of? Describe how their approach was both similar and different to your approach.

3. Did the group brainstorm how to solve each problem, or was another approach taken, such as working independently and comparing solutions? Did every group member agree on this overall approach? If not, how did the group reconcile these different preferences?

Figure 5.12: CBE 201: Reflection Question Set 1 Following Each Project [168].

5.4.3 Challenges

Initially, the intent was to attempt incorporation of these concepts throughout the entire curriculum (in at least one course every semester – see the program structure in Fig. 5.11) in each academic year. This has presented several challenges, aside from the extremely technical course content and course structure. Broad incorporation requires significant coordination and collaboration with and between faculty members within the CBE department to determine the best approach for each course and avoid redundancy across courses that may result in decreased student participation. In addition, the activities need to be relatively self-sustaining in the course once integrated since a required follow-up for every initiative in every course over time is unachievable, given the time expense. This poses multiple challenges and will continue to be an ongoing process. Additionally, individual faculty must not only recognize the value of addressing these topics but must also be willing to make adjustments within their course structure to accommodate these topics in some form.These challenges will be discussed in more depth in following sections. Overall, however, the challenges faced and lessons learned from incorporating DEI into these courses, including maintaining these activities in the courses as well as engaging and motivating faculty, are informative for in future planning and implementation initiatives as the department makes steps to become more inclusive.

5.5 DEI in Engineering Education: The Current Landscape

In Section 5.4, previous work related to incorporating DEI into the CBE curriculum in the first and second years is discussed. There are, however, numerous approaches to incorporating these topics into engineering curriculum. Some of these approaches are direct and forthcoming with their intent, related to course content. Others, however, are more broad and provide a general overview, or are indirect and focused on pedagogy to create a more engaging, accessible learning experience. Many of these approaches are more directly related to pedagogy, such as altering class structure to include flipped classroom environments and increased student involvement throughout the class sessions. There have been many recent works with a wide variety of approaches to address DEI in the field of engineering education.

5.5.1 DEI in Course Curriculum

Some recent research works have primarily approached diversity through the lens of engineering curriculum and focused on modifying the course curriculum [173, 184, 185, 188, 193, 212, 215, 216]. For example, two DEI activities have been included in undergraduate chemistry courses at Gulf Coast State College in Nakamura (2002) [193]: reflection questions after attending diversity and inclusion community engagement talks, and completion of the IAT and a subsequent short reflective essay. As mentioned in Sec. 5.4, although there is value in recognizing the presence of implicit biases in all interpersonal interactions and the potential impact of these biases, as well as the limited number of tools to be able to measure and quantify this, growing research suggests the IAT has significant flaws and may not in fact be measuring what is intended [208–211].

In a significantly different approach, Wendland and Worthington (2023) [188] modified an upper-level biomedical engineering course to dedicate class time to the incorporation improv activities, as well as conducted pre- and post-implementation surveys to gauge student engagement, self-confidence, communication, and comfortability in the classroom. Bansal, *et al.* (2023) [185], though, details numerous approaches for integrating DEI into engineering curricula to provide guidance to faculty and instructors, given the new ABET DEI criteria in biomedical engineering curricula. The importance of faculty training in this area is stressed as a jumping off point in the process of DEI incorporation into the curricula. Beyond this, two primary actionable recommendations are to revise language to be more inclusive, particularly in relation to disabilities, and to take surveys to assess the student experience in the program and the effectiveness of teaching interventions. Many of the other suggestions in this work relate specifically to biomedical engineering design courses; however, they do provide some examples and thought exercises in this arena that could be beneficial for reflections in related technical courses regarding the impact of engineering designs that do not account for diversity.

Others have approached diversity through a combination that includes both engineering curriculum and course content using a more well-rounded, holistic approach that looks at the course as a whole instead of just considering the incorporation of individual activities [173, 184].

For example, Rice and Mays (2022) discuss the integration of DEI through the modification of the course structure of a fluid mechanics course at the University of Colorado Denver (CU Denver) [173]. Although implemented for fluid mechanics course for civil-engineering majors, the approach is directly applicable to equivalent courses in other engineering majors. These changes were originally made accommodate a hybrid online learning course with both asynchronous and synchronous components as a result of the Covid-19 pandemic; however, they can be easily adapted for an in person course. They implemented numerous changes while taking into consideration the theoretical framework of Culturally Relevant Pedagogy [217]. First, they revised the syllabus language and phrasing to be more inclusive and have a more supportive, less authoritative tone while still setting the same standards and requirements for the course; additionally, they included a land acknowledgement statement on the syllabus, which is quite applicable given that CSU is a land-grant institution. In modifying the course content, they asked themselves two questions:

- "Do these course materials marginalize anyone?"
- "Do these course materials represent a missed opportunity to engage anyone?"

While doing this evaluation of the course content, they also replaced some of the personal nomenclature throughout the course material with technical nomenclature. For example, changing "Bernoulli" to "Energy Equation" and "Buckingham Π" to "Dimensional Analysis" so that the more descriptive wording that is used removes the implication that fluid mechanics is "exclusively by and for white men" [173]. They also attempted to change at least some of the in-class examples to be more engaging by adding cultural context (since this is a time consuming challenge) as well as changed the videos presented throughout the course, replacing videos with homogeneous demographics which students may have difficulty looking beyond, with videos that address the same topics with a more accessible presentation. In this work, they also provide suggestions for engineering faculty looking to infuse DEI into engineering courses. Although utilizing this approach to evaluate those two questions for all course materials in a specific course may be incredibly time consuming overall, it is one that can be applied systematically and incrementally over time to all courses within the curriculum.

In their approach, Busch-Vishniac and Jarosz (2004) [184] developed six steps to follow when evaluating revisions to engineering curriculum. Unlike the work of Rice and Mays (2022) [173], their approach is geared more generally towards the entire engineering curriculum in a particular program and not just one specific course. In their approach, they suggest adapting courses starting from the bottom up, making the first year or two and earlier engineering courses the students encounter more integrative and interdisciplinary. They also outline what numerous different universities have attempted to do in this area since the early 1990's to determine what has previously been done to see if it can be built upon (Step 1).

In addition, they discuss the lack of social relevance in many engineering programs in spite of the critical need for its inclusion given the overall fundamental purpose of engineering and its relation to the needs of society. The poignant lack of recognition of female and minority scientists, engineers, and physicists in many textbooks as compared to males, including the use of their work and a lack of acknowledgement by full names, and the need for improving diversity and multiculturalism in the curriculum is also discussed. Since this can result in overly crowded curricula, tough discussions regarding the necessity of all the current inclusions and prioritization may be required to streamline the curriculum instead of extending the time to graduation. This is particularly true since this is an area where engineering departments already struggle, with most allowing for little flexibility for students to be able to achieve graduation in four years, which is more likely to negatively impact students from historically excluded or marginalized groups.

On a related front is the topic of required prerequisites, and which are actually critical for the degree and which, although they may be beneficial, are not actually a necessity. They also suggest departments consider reorganizing parts of the course sequence, if need be. Another noteworthy topic they touch on are the assumptions many faculty make regarding the level of ability of the entering students, which can prove to be an impediment for many minority students since not all high schools offer the same material or the same quality of education. Although this work was published well before the Covid-19 pandemic, this topic seems especially relevant for current students since, from personal experience teaching first-year engineering students before, during, and "after" the pandemic officially ended, the distribution of preparedness seems to have broadened as a result. Moreover, in line with what the CBE department has already been working towards, is defining the engineering culture as one of collaboration instead of competition to enhance academic experience, decrease isolation of the students, and improve teamwork and the overall well-being of the students. In addition, this work addresses multiple other areas worth considering to address issues of accessibility in engineering programs.

Shields (2023) [218] builds upon some of the areas suggested for improvement regarding DEI in engineering education in Busch-Vishniac and Jarosz (2004) as they outline their approach in a pilot program at the University of Pennsylvania to incorporate justice, equity, diversity, and inclusion (JEDI) material into a core introductory bioengineering course, following an "ethics across the curriculum" pedagogical approach [219, 220]. In this approach, they embed content related to ethics in the course alongside the technical content to instil a sense of ethical responsibility as well as provide societal relevance. Their goals included having students learn about the intersectionality of bioengineering practice with JEDI through an interdisciplinary context in addition to having students contribute to and participate in an inclusive environment while learning about and discussing these topics. Although they consider this to be a pedagogical approach, it seems to be more of a combination approach related to both pedagogy and content. Many of the concepts and topics they introduced into the program are relevant to and can be easily incorporated into the current CBE program, even in a piece-wise fashion.

5.5.2 DEI Primarily Through Pedagogy

Instead of focusing on curriculum, other works have taken a more pedagogical approach to DEI in engineering [174, 181, 218, 221]. In Krestra (2021) [181], for example, the skills required for innovation, creativity, and reliability in engineering are framed and discussed in the context of topics related to DEI. Although the work is less focused than many others and does not provide much in the way of substantial examples to put into practice, it does provide some ideas on active learning and collaborative activities to incorporate into a classroom setting.

A wide range of specific activities and interventions that can be implemented in STEM courses are introduced in Casad, *et al.* (2018) [215]. A review of social-psychological interventions that have been incorporated in education as well as their outcomes as related to grade point averages (GPAs) was conducted by Yeager and Walton (2011) [222]. Tibbetts, *et al.* (2016) [216] describes social-psychological interventions in undergraduate biology with a goal of promoting positive academic outcomes for underrepresented groups, some of which were implemented in a class setting.

Although this work is specifically related to biology courses, they could be just as easily incorporated in engineering courses. While this is a pedagogical strategy that could potentially benefit all students, this research suggests a larger positive impact for students from historically-excluded populations.

Farrell, Godwin, and Riley (2021) [174] approach inclusive teaching practices in engineering through a lens of a sociocultural learning framework which emphasizes the importance of collaborative dialogue with peers and larger context in an individual's learning development and learning environment. They discuss some exclusionary aspects of the engineering culture that act as barriers to both participation and inclusion. Additionally, they provide some strategies as a jumping off point for incorporating DEI into chemical engineering curriculum, with a focus on learner identity and classroom context. These strategies include an active approach to an inclusive classroom in which the diversity of students' identities are valued in their instructor interactions instead of an approach that assumes an inclusive classroom is one that lacks bias and has equitable treatment for all students. Examples of this in practice are using terms like "everyone" instead of "you guys" as well as using students preferred names and pronouns. Other strategies include the importance of recognizing student-wellness, using an asset-based model that approaches diversity from a positive view of unique strengths and perspectives these students bring to the field versus the view of what they lack, setting classroom norms, and using socially relevant examples.

Butterfield, McCormick, and Farrell (2018) [221] discussed some of the same strategies mentioned in Farrell, Godwin, and Riley (2021) [174] regarding the inclusive classroom and students' identities, and also mentioned assisting in access to support networks and being visible as an ally.

5.5.3 Transformative Approaches to DEI in Engineering

Looking beyond just incorporating DEI in current course structures, pedagogy, and curriculum, some take an ambitious, more transformative approach to DEI in engineering education that involves the addition of new courses, entirely reworking the program, or significant restructuring of current courses. For example, the beneficial impacts of experiential learning in engineering education are discussed numerous recent works, with the focus of many related to biomedical engineering education [223–232], while others discuss the benefit of incorporating immersive research experiences for undergraduates (REU) into engineering programs [233].

Nease (2021), however, discusses concrete approaches for modifying the undergraduate engineering curriculum through community-engaged design initiatives (also known as service-learning) [194]. In Nease (2021), undergraduate participation in these is discussed to help students develop greater awareness of the complexities of real world engineering problems while also developing more empathetic understanding and ethical, social, environmental, and economic issues that may be encountered. This work also discusses a major transformation to the entire first-year engineering curriculum by combining four courses into one large course at McMaster University to employ an experiential framework that utilizes a series of five community-engaged design projects (in teams) along with subsequent corresponding reflections on each experience. This work also discusses the addition of in-class workshops, open discussions, and discovery-based learning to chemical engineering courses throughout the following years in the curriculum. As a way to elicit an individual students understanding and mastery of the course material, the concept and benefits of collaborative assessments and collaborative re-testing are discussed, along with a step-by-step example process of how this is put into practice in an engineering course, as well as an evaluation of the results of these practices, including student surveys and grade distributions. While these practices don't necessarily directly relate to adding DEI in the course content, they are related to ways DEI can be utilized to alter the pedagogy for some engineering topics to make the abstract concepts more clear, improve teamwork and understanding from fellow students, and increase the accessibility of the material to all students, not just students that learn in one particular way.

Reynante (2022) [195] is another work that also discusses the addition of community-engaged learning initiatives in an engineering program at a university on the West coast, with a "design-for-justice" approach as opposed to a "design-for-charity" approach that others often take. The "design-for-justice" approach incorporates participatory community involvement, considers reflexivity to be critical, and utilizes asset-based community framing to focus on causes of inequity instead of symptoms of inequity. In this program, two courses are required, an introductory course and laboratory course (with a field practicum), out of several courses taught by engineering faculty in the university that infuse social justice and service-learning. This case study is conducted for one of the particular projects this community-engaged engineering program worked on, which focused on a lack of affordable and reliable lighting in a village in the Philippines, which began when the program was contacted by a non-governmental organization (NGO) in the country that aims to alleviate poverty. This study was guided by a social constructivist framework, which considers learning to be an active, social process in which conceptual understanding is built by utilizing previous experiences, knowledge, interactions with specific context and culture, and reflections of those [234–236]. In this work, Reyante provides a thorough example of positionality by evaluating his positionality in reference to the current study. In evaluating the study, he used a deductive qualitative analysis approach, coding, and thematic analysis to investigate if/how participants demonstrated shifts in mindsets.

A challenge of many of these approaches is the requirement for substantial resources, both in terms of time and financial capital. Additionally, these more transformative approaches require substantial engagement and collaboration with individuals, entities, and businesses outside of the university. Without these, it may still be possible to implement some community-engaged or real-world initiatives on a smaller scale that is more feasible, given the current resources of the department.

5.5.4 A Broad View of DEI in Engineering Education

There are also some works that are primarily focused more on the big picture of DEI in engineering education [179, 180, 182, 183, 186, 212–214, 237–246], including what works overall and what doesn't, what are some challenges that may be encountered, how students with various identities and minoritized statuses perceive these efforts, the impact of level of engagement of faculty, etc.

The evaluation of "successful" science and engineering programs for undergraduate women as measured by an outcome of the number of undergraduate degrees awarded to women in science and engineering was completed by Fox, Sonnert, and Nikiforova (2009) [179]. In these programs, they found the most successful programs were the ones in which the issue being addressed was defined and approached from a more comprehensive, institutional/structural perspective instead of a more individualistic perspective. For example, while all of the programs were ultimately working on recruitment and retention struggles, more successful programs defined the overall issue as a combination of underrepresentation and minoritized status, grading systems to "weed-out" students, faculty and classroom biases, and a lack of a support pipeline as well as stressed a need for creating an equitable environment. The least successful, on the other hand, had a more simplistic view of the overall issue, such as recruitment and retention (i.e., the number of women participating and staying) as related to personal or gender deficiencies, and these programs had a goal of attracting, recruiting, and retaining women by increasing "outreach" and mentoring. While these are not bad or useless strategies and they can,in fact, be quite beneficial and successful [180], they can't be the only strategies or initiatives taken; they can't be completed in a vacuum, with the absence of more substantial changes that take into consideration the "why" of both recruitment and retention issues, including the overall environment of the institution and/or program itself.

Equity in engineering classrooms is addressed in Secules and Masta (2020) [182], with a discussion of the marginalizing experiences of students in underrepresented and historically excluded groups, such as in relation to race, sexual orientation, first generation status, socioeconomic status, gender, as well as gender non-conformity, and ability. In this realm, they also discuss how many studies focus on only a single dimension in relation to this, with a primary focus often being binary gender despite the status of many students in multiple marginalized groups simultaneously. Instead of directly addressing ways to implement culturally relevant pedagogy into the curriculum like Rice and Mays (2022), this work, in contrast, broadly discusses the challenges present in the creation of culturally responsive pedagogy (CRP), covered in depth in Sleeter (2012) [183], particularly the risks of *essentializing* CRP by assuming a fixed characterization of cultures or students belonging to a particular group, taking on a simplistic view or *trivialization* of what CRP is, implementing only superficial pedagogical changes as a result, and a failure in connecting its use with improving student learning and achievement. This work also discusses the unintended negative impact meritocracy and competition in engineering courses can have on some students in a class environment. In examining changes to curriculum, Secules and Masta suggest considering a series of questions, which include:

"What does it mean to imagine an engineering curriculum that is not steeped in whiteness and masculinity? Whose culture should we respond to and how?"

"Can all students be supported in learning and growth as engineers without some students becoming constructed as failure in comparison to others?"

"How does the curriculum of engineering education pursue liberatory, critical, and justice-oriented practice within subject matter traditionally nested within neoliberal capitalism?"

In doing this, they suggest multiple areas within a course to consider where equity issues may be encountered, including curriculum, course texts, communication, grading, interactions, and support. They also discuss the benefits of an expansion of faculty learning to better understand their individual "learning edges" and positionality in relation to their students as well as diversity, equity, and inclusion.

Valdez and Kelp (2023) evaluate student perceptions of inclusive pedagogy in undergraduate STEM classrooms, particularly students in historically marginalized and underrepresented populations [242]. They, along with others [212–214, 237–241], also discuss the relationship between aspects such as grit, self-efficacy, self confidence, science communication skills, and science identity and undergraduate STEM students' personal motivations, community engagement, retention in STEM programs, and career attainment. These works collectively provide valuable information regarding motivations for DEI interventions and areas to consider when designing these.

Thoman, *et al.* (2021) [186] address some of the challenges in achieving the integration of DEI initiatives and interventions into undergraduate classrooms, despite the relative ease and evidence-based success of many of them, which they describe as a "resistance to action model". Additionally, this and many other works discuss some of the research that has been done over decades that show that ignoring underrepresentation, including gender and race to take a color-blind approach, can in fact exacerbate inequality, undermine interracial interactions and diversity efforts, and reduce sensitivity to racism [247–252]. Collopy, *et al.* (2022) [187] discuss challenges regarding faculty hesitancy and resistance as well as workshops conducted as part of a program with approximately 20 faculty members in multiple stages of their career that represent every department in the College of Engineering, Design, and Computing at CU Denver. This program, called the Faculty Learning Community (FLC), follows the Colorado Equity Toolkit three-component framework [253]. Park, *et al.* (2020) discuss what engaged faculty looks like as related to DEI, and address approaches that seem to work in efforts to engage faculty in incorporating diversity-related content in their courses [243].

Hartman, *et al.* (2019) [244] also address the need for an inclusive environment, pedagogy, and curriculum changes, and framed their work in constructivist learning theory [236], like Reynante (2022) [195], and critical (liberatory) pedagogy [254], which orients education towards social justice and the liberating of students from oppression. In addition, they discussed student versus faculty perceptions, the inclusion of faculty workshops regarding DEI and curriculum changes, the benefit of mentoring for first-year and transfer students, conducted a number of surveys, and touched on admissions standards. The misalignment in stated goals for the incorporation of DEI into undergraduate engineering programs (particularly as related to women in engineering) and the actual implementation approaches taken are discussed in Fox, *et al.* (2011) [246]. This work stresses the importance of consideration of the institutional, college, and program environments while evaluating various DEI inclusions and implementation, as well as the importance of potentially larger scale activities regarding faculty and program / classroom climate. Lezotte (2023) also looks at the institutional context and environment, and how it shapes diversity-related efforts within the university [245].

Overall, the expansive field of recent work in this area provides ideas and examples for incorporation of DEI initiatives into the pedagogy, curriculum, or a combination of approaches as well as many larger-scale transformative approaches to adapting the curriculum. Additionally, there is a wide array of information regarding the perceptions, motivations, and sense of belonging of students in underrepresented and historically excluded groups, and the broader view of the impact of various interventions and activities to help direct engineering departments in their approach.

5.6 Moving Forward: Possibilities for the CBE Department

There are numerous opportunities for the CBE department at CSU to continue to incorporate DEI into the CBE curriculum and program as a whole, some of which are proposed to extend efforts that have previous been done (5.4) and some of which have not been previously explored. These efforts are proposed to help the department cultivate a more supportive, inclusive educational environment and culture that better reflects the diversity within society as a whole as well as helps improve the retention and success underrepresented and historically excluded students within the CBE and dual degree CBE/BME programs. These initiatives also have the potential to benefit all CBE students in achieving success in their academic career and in their future professional careers by helping them foster inclusive professional identities and improve their ability to work well in diverse teams.

5.6.1 Proposal: Adapting for Diversity Growth and a More Inclusive Future

Although I propose that a substantial approach to the incorporation of DEI in the CBE department which includes the systematic and strategic implementation of wide-spread revisions in the curriculum and program overall would be the most effective and impactful for the CBE department at CSU, this could be applied over time starting from the bottom up. The proposed approach involves changing the tone broadly throughout the program to increase cultural competence and inclusive pedagogy. In this approach, the department overall could utilize some of the same tools that can also be incorporated on an individual course level, with a goal of creating a more inclusive educational environment overall that addresses some historical disparities that continue to exist and proliferate in the entire engineering field.

I propose that, initially, the incorporation of DEI in the CBE department be integrated throughout all of the courses in the CBE curriculum by utilizing an approach to includes aspects of the strategies discussed in Busch-Vishniac and Jarosz (2004) [184] and described in Rice and Mays (2022) [173], which have been previously implemented in an engineering course at CU Denver. However, instead of each individual faculty member looking at only a specific course or the course they "usually" teach, I believe a more effective approach is one which is systematic and strategic, that also has the potential to add an element of consistency regarding departmental DEI values and goals throughout the program. Of course, for this approach to be effective, the department needs to have some collective understanding and consensus of what its core values are as well as its DEI values and goals are. Some of the questions that may need to be answered by the faculty both individually and overall as a group are:

- Why do you think enrollment has not recovered from the pandemic? What is the problem?
- What do you see as driving forces for the low enrollment of historically marginalized populations?
- What do you see as main issues that need to be addressed in the department in relation to enrollment, DEI, retention, departmental culture, etc.?
- Are there any areas that can be identified as generally in need of changes or improvements?
- Do you have ideas for changes or improvements that could be made that you believe would be beneficial to remedy these?
- What are the collective DEI values for the department? What are the ultimate goals?
- Where do you envision the department in 2 years? What about 5 years?
- What do you see as a primary roadblock in achieving this?
- What do you foresee as barriers to incorporating DEI in the department and the program?

These questions could be addressed in a facilitated workshop in which faculty learn and think together about DEI in both engineering education and an institutional context. Given the known challenges regarding the range of faculty comfort with DEI topics [182, 183, 186, 243, 245], additional faculty training and workshops may also be extremely beneficial to help faculty understand these issues better in order to consider them in the courses, but also to understand their own positionality in terms of understanding DEI issues in the classroom and department [182, 183]. Collopy, *et al.* (2022) [187] provides valuable insight regarding an approach to increase faculty engagement and investment in DEI. Additionally, the three-component framework of the Colorado Equity Toolkit [253], which includes resources for self inquiry, course design, and creating community, is specifically designed for educators, including instructors and faculty, in the State of Colorado. According to the Colorado Department of Higher Education:

> *"Although Colorado enjoys one of the highest education attainment rates nationally, the share of the white majority population that has earned a credential is more than twice that of our fastest-growing group—Latinx and Hispanic Coloradans—and more than 1.5 times the share of African American, Black and American Indian Coloradans.*
> *We can and must do better. As higher education leaders, it's on us to foster welcoming, thriving learning environments for all—especially in the classroom." [253]*

With this in mind, I suggest that the DEI Committee brainstorm to not only answer these questions, but also develop a list of what aspects of DEI they believe should be consistent throughout the program (and department overall). It may also be beneficial or necessary to ask external advisors about the DEI skills graduates will need. These answers and this list can be used as a "jumping in" point to initiate a discussion within the faculty and staff regarding DEI since, as many research works discussed in Sec. 5.5 have noted, effective, broad implementation of DEI initiatives requires not only cooperation of the faculty, but engaged faculty members.

The faculty and staff should identify any specific aspects of DEI the department wants to unite behind and emphasize for the students. For example, if the faculty and staff able to reach consensus and get "buy-in" from members, the department could add some "universal" changes throughout the program to *all* syllabi, including: the addition of a land acknowledgement, a statement regarding respect and inclusivity in the classroom, and the incorporation of general language changes to be more inclusive and supportive on syllabi. This would help affirm for students that the values and culture the department strives for both overall and in their individual courses is indeed recognized as important and supported by all faculty.

Additionally, I recommend that *every* course in the curriculum replace some of the personal nomenclature with more technical, descriptive terminology in the course material. For example, using For example, using "Dimensional Analysis" instead of "Buckingham Π" in course materials. It is my understanding that the basic content of courses is reviewed and discussed in faculty meetings, particularly to ensure that each course is satisfying the requirements of ABET. In this process, each course review could also address potential DEI issues and the consideration of equity in course structure and design, as well evaluating the two questions suggested by Rice and Mays (2022):

- "Do these course materials marginalize anyone?"
- "Do these course materials represent a missed opportunity to engage anyone?"

Further, in the review of courses, faculty together could modify language in current assignments and problems to be more inclusive as well as work to incorporate at least a couple of example problems that include cultural context and real-world relevance. Drawing on the collective breadth of faculty experience and knowledge may, in fact, be more time efficient than if it were done by an individual faculty member since some faculty members may have more to offer (based on background and knowledge) regarding these aspects for particular classes, even if they do not frequently teach them. For upper level courses, I also suggest the addition of case studies which demonstrate the importance of diversity in engineering design, as well as the incorporation of concepts such as sustainability and ethical sourcing in engineering design course.

5.6.2 A Next Step: Course-Specific Interventions

For a small number of courses, I propose some specific DEI inclusions, particularly in lower level courses. For example, if the first year seminar course (CBE 180A2) that was taught in Spring 2023 is continued, I recommend the addition of discussions in that course as well as CBE 101 regarding the benefits of inclusive language. If the science communication workshop that was incorporated in CBE 101 has not been continued from Spring 2023, perhaps it could be included in this first year seminar because, in addition to continuing the science communication workshop, I suggest continuing the efforts and building on it in following years. To extend activities related to science communication into the later years in the academic program, Kelp and Hubbard (2021) [255] describes curriculum for both 200-level and 400-level science communication courses, parts of which could be effectively incorporated into CBE courses. For the first-year seminar, some additional possible approaches can include (but aren't limited to): assignments on recognizing and understanding the impact of microaggressions, and shedding light on challenges that affect women and/or underrepresented and historically excluded populations, such as pay discrepancies or other inequities in industry.

5.6.3 Taking It Further: Transformative Additions

Beyond the proposed adaptations of the CBE program discussed in Sec. 5.6.1, I propose that the CBE department (and every engineering department, in fact) should consider the addition of at least one non-traditional elective course that incorporates experiential, service learning in the form of community-engaged sustainable design initiatives or projects. Whether these projects are geared toward the local community or other communities, possibly with a focus on disadvantaged or struggling communities with a current lack of resources or access, this course could greatly increase undergraduate students' awareness of the extensive complexities and challenges involved in the process of engineering solutions in a real world setting, including the difficult balance that must be achieved regarding ethical, social, environmental, economic, and political issues. These projects would not only help students with engineering skills, but would also help build students' skills in communicating with others with different backgrounds and education levels, strengthen their team work, and potentially increase their empathy for others. This plan is a reduced, smaller scale version of the EPICS program at Purdue University [256] or the vertically integrated projects (VIP) that can span for years at universities such as Georgia Tech and NYU Tandon School of Engineering [257–259]. If CBE is able to achieve this, it could be used as a model for other courses as well as for other departments in the future.

In contrast to having each group work on the same project or a couple of different projects through the semester, another possible approach I envision for this course structure includes groups rotating through different projects in a sequence. Depending on if problem formulation is included, this course would be divided into either three or four steps, so the semester is broken up into either thirds or quarters. If problem formulation is involved, there would be a Step 0 before the following three steps:

1. Identify and consider potential sustainable engineering solutions for an identified problem, issue, or opportunity (working with community or company contacts or collaborators, if necessary). Flesh out a sustainable engineering design for the selected solution;
2. Switch projects. Evaluate the designs provided by groups for a different project in Step 1 to identify any potential issues or concerns with the proposed designs. If multiple designs are evaluated, select the best design based on the options or potentially develop a hybrid approach. Modify or revise the design to resolve the problems or issues identified;
3. Switch projects. Take the revised, improved design for a separate in Step 2 and create a plan for implementation (working with community or company contacts or collaborators, if necessary) potentially including a cost analysis, analysis of local challenges, sourcing issues for materials, etc. that may be encountered.

This course could include a set of 4-6 project groups, each with 4-5 students from second-year and above. Two or three separate projects or design initiatives could be considered by the different groups during this course. This rotation strategy would not only include the science communication and mentorship aspects that are present in the approach in which groups only evaluate a single project from start to finish, but adds additional aspects of solving engineering problems that they all may encounter in real-world situations, the complexities of communicating a design to others outside of the team as well as the critique of others' work, which includes the recognition of potential issues with and the need to improve or modify others' designs. Additionally, in either of these approaches, I suggest each group have a "faculty mentor" beyond just the course instructor that they can bounce ideas off and ask for suggestions.

Of course, given the ambitious nature of this recommendation, there is an understanding that challenges are present in attempting to achieve this, including the substantial investment of time, the need for significant collaboration with the communities involved, the need for collaborative faculty involvement, and a need for funding. However, there are funding opportunities for endeavors like this through the National Science Foundation (NSF) for both faculty and for graduate students who are interested in being involved.

Another possibility is, instead of adding a course, to have a community-engaged project option in Chemical and Biological Engineering Design I and II (Senior Design). However, the restriction to only higher level students is a disadvantage. The benefit of a separate elective course is the possibility for students in years two through five to take this course, with strategic placement of students in project groups to account for the variability of skill level as a result of their current education level. Not only could this course be an excellent learning opportunity for all students, but it could also be a built-in opportunity for upper level students to provide guidance and mentoring for others, but also for all participating students to gain skills and confidence in science communication.

5.6.4 Summary

Overall, an in depth look at the current demographics, persistence, and graduation rates of the Walter Scott, Jr. College of Engineering at CSU 5.3.1 along with a more in depth evaluation of the CBE department 5.3.2 is completed here 5.3. A wide range of different approaches to both understanding and incorporating diversity, equity, and inclusion in undergraduate engineering education are explored in this work that could potentially benefit the CBE department in its efforts to increase diversity and inclusion. Additionally, a variety of approaches have already been incorporated in the first- and second-years courses in the CBE program over the last six years. Through lessons learned in this process as well as the current body of research regarding the challenges regarding DEI in engineering, possibilities for implementation, and evaluation of the effectiveness of many of these, a more substantial, holistic approach to DEI is proposed here as a viable, yet achievable option for the CBE department. Additionally, an ambitious approach for an additional elective in the course is also suggested. Together, incorporating these could be a large step forward for the department to both create a more inclusive, supportive educational experience for undergraduate students and better enable them to handle the complex challenges they may face in their careers as a result of limited resources and our diverse, increasingly interconnected world.

Chapter 6

Future Work

In this work, numerous computational approaches for analyzing biological systems were explained and demonstrated. This demonstration included addressing differences between the computational methods as well as challenges in utilizing these methods in the analysis of real data. As related to challenges, further investigation is needed in the form of both additional experimental data and potentially a revised modeling approach to improve the understanding of the interplay between temperature, ionic strength, and the level of supercoiling of the DNA molecule on the conformational dynamics of the DNA double-helix and subsequent oligonucleotide probe binding dynamics. Further analysis is also necessary to better elucidate the model as related to the intuitive physical understanding of the system.

Many of the concepts discussed in Chapter 5 regarding DEI in undergraduate education, as well as and the proposal for increasing DEI in undergraduate education, are applicable to engineering programs throughout higher education, not only the program at CSU. When considering other engineering programs, however, it is crucial to keep in mind that each engineering program, department, and college are unique, with their own particular strengths and weaknesses. Given this, it is important to consider the individual culture and characteristics of each individual department, including the size and demographics of the undergraduate population. Additionally, it is important to consider each department's core values, educational mission, and any challenges or circumstances that are specific to them when considering why and how to implement DEI.

Regarding DEI in the undergraduate education experience in the CBE department at CSU, the demographic analysis suggests a need for change throughout this department to create a more inclusive, supportive educational experience for undergraduate students. This change would also potentially increase both enrollment and retention of a more diverse undergraduate student population. However, in order for this to materialize, the faculty and staff would need to collectively evaluate whether and in what ways their current educational environment is aligned or misaligned with its mission and goals related to the undergraduate educational experience and DEI. A proposal for moving forward and increasing DEI throughout the undergraduate program is provided. Additionally, an ambitious approach for an additional elective in the curriculum related to a project-oriented, community-engaged, experiential learning course is suggested. Together, incorporating these could be a large step forward for the department to both foster a more inclusive environment for undergraduate students and better enable these students to handle the complex challenges they may face in their careers with limited resources in a diverse, interconnected world.

Beyond just the undergraduate educational experience, future work for many universities includes a thorough analysis overall, including the culture and support structure in their graduate program as well as for their faculty/staff. If there are issues with diversity, equity, and inclusion in these areas, it would be unrealistic to assume this not be felt by the undergraduate student population. For an engineering department to truly be prepared to grow and improve, it needs to have engaged, motivated leadership and a willingness and investment in fostering a more supportive, inclusive culture throughout the department, not just for the undergraduate students.

References

[1] G Neuert, B Munsky, R Z Tan, L Teytelman, M Khammash, and A van Oudenaarden. Systematic Identification of Signal-Activated Stochastic Gene Regulation. *Science*, 339(6119):584–587, January 2013.

[2] Adrien Senecal, Brian Munsky, Florence Proux, Nathalie Ly, Floriane E Braye, Christophe Zimmer, Florian Mueller, and Xavier Darzacq. Transcription factors modulate c-Fos transcriptional bursts. *Cell Reports*, 8(1):75–83, July 2014.

[3] Shane Scott, Zhi Ming Xu, Fedor Kouzine, Daniel J Berard, Cynthia Shaheen, Barbara Gravel, Laura Saunders, Alexander Hofkirchner, Catherine Leroux, Jill Laurin, David Levens, Craig J Benham, and Sabrina R Leslie. Visualizing structure-mediated interactions in supercoiled DNA molecules. *Nucleic Acids Research*, 46(9):4622–4631, 04 2018.

[4] Shane Scott, Cynthia Shaheen, Brendon McGuinness, Kimberly Metera, Fedor Kouzine, David Levens, Craig J Benham, and Sabrina Leslie. Single-molecule visualization of the effects of ionic strength and crowding on structure-mediated interactions in supercoiled DNA molecules. *Nucleic Acids Research*, 47(12):6360–6368, 05 2019.

[5] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[6] Matthew H. Spitzer and Garry P. Nolan. Mass cytometry: Single cells, many features. *Cell*, 165(4):780 – 791, 2016.

[7] Charles Gawad, Winston Koh, and Stephen R. Quake. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, 17:175–188, 2016.

[8] Howard M Shapiro. *Practical Flow Cytometry*. John Wiley & Sons, 2005.

[9] Tatsuya Morisaki, Kenneth Lyon, Keith F DeLuca, Jennifer G DeLuca, Brian P English, Zhengjian Zhang, Luke D Lavis, Jonathan B Grimm, Sarada Viswanathan, Loren L Looger,

and *et al.* Real-time quantification of single rna translation dynamics in living cells. *Science*, 352(6292):1425–1429, 2016.

[10] Arjun Raj, Patrick Van Den Bogaard, Scott A Rifkin, Alexander Van Oudenaarden, and Sanjay Tyagi. Imaging individual mrna molecules using multiple singly-labeled probes. *Nature Methods*, 5(10):877–879, 2008.

[11] A M Femino, F S Fay, K Fogarty, and R H Singer. Visualization of Single RNA Transcripts in Situ. *Science*, 280(5363):585–590, April 1998.

[12] Thomas A. Kunkel and Katarzyna Bebenek. DNA replication fidelity. *Annual Review of Biochemistry*, 69(1):497–529, 2000. PMID: 10966467.

[13] Janine N. Post, Sandra Loerakker, Roeland M.H. Merks, and Aurélie Carlier. Implementing computational modeling in tissue engineering: Where disciplines meet. *Tissue Engineering Part A*, 28(11-12):542–554, 2022. PMID: 35345902.

[14] Daniel Pearce, Sarah Fischer, Fatama Huda, and Ali Vahdati. Applications of computer modeling and simulation in cartilage tissue engineering. *Tissue Engineering and Regenerative Medicine*, 17:1–13, 2020.

[15] Md. Shakhawath Hossain, D.J. Bergstrom, and X.B. Chen. A mathematical model and computational framework for three-dimensional chondrocyte cell growth in a porous tissue scaffold placed inside a bi-directional flow perfusion bioreactor. *Biotechnology and Bioengineering*, 112(12):2601–2610, 2015.

[16] R Bardini and S Di Carlo. Computational modeling and optimization of biofabrication in tissue engineering and regenerative medicine - a literature review. *bioRxiv*, 2023.

[17] T.A. Brooks, S. Kendrick, and L. Hurley. Making sense of g-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, 277:3459–69, 2010.

[18] A.C. Ketron and N. Osheroff. Topoisomerases and cancer. In L.I. Grossman, D.A. Julin, J.M. Kaguni, and I.R. Lehman, editors, *Molecular Life Sciences*, pages 1–9. Springer Science + Business Media, New York, 2014.

[19] Xue Lei, Wei Tian, Hongyuan Zhu, Tianqi Chen, and Ping Ao. Biological sources of intrinsic and extrinsic noise in ci expression of lysogenic phage lambda. *Scientific Reports*, 5(1):13597, 2015.

[20] Jonathan M. Raser and Erin K. O'Shea. Lnoise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, 2005.

[21] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[22] Brian Munsky, Guoliang Li, Zachary R. Fox, Douglas P. Shepherd, and Gregor Neuert. Distribution shapes govern the discovery of predictive models for gene regulation. *Proceedings of the National Academy of Sciences*, 115(29):7533–7538, 2018.

[23] Benjamin D Maier, Luis U Aguilera, Sven Sahle, Pascal Mutz, Priyata Kalra, Christopher Dächert, Ralf Bartenschlager, Marco Binder, and Ursula Kummer. Stochastic dynamics of type-i interferon responses. *PLOS Computational Biology*, 18(10):e1010623, 2022.

[24] Md Rifat Hasan, Ahad Amer Alsaiari, Burhan Zain Fakhurji, Mohammad Habibur Rahman Molla, Amer H Asseri, Md Afsar Ahmed Sumon, Moon Nyeo Park, Foysal Ahammad, and Bonglee Kim. Application of mathematical modeling and computational tools in the modern drug design and development process. *Molecules*, 27(13):4169, 2022.

[25] Didier Gonze, Katharine Z Coyte, Leo Lahti, and Karoline Faust. Microbial communities as dynamical systems. *Current opinion in microbiology*, 44:41–49, 2018.

[26] Cynthia Shaheen, Cameron Hastie, Kimberly Metera, Shane Scott, Zhi Zhang, Sitong Chen, Gracia Gu, Lisa Weber, Brian Munsky, Fedor Kouzine, David Levens, Craig Benham, and

Sabrina Leslie. Non-equilibrium structural dynamics of supercoiled DNA plasmids exhibits asymmetrical relaxation. *Nucleic Acids Research*, 50(5):2754–2764, 02 2022.

[27] Richard M. Fye and Craig J. Benham. Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA. *Phys. Rev. E*, 59:3408–3426, Mar 1999.

[28] James M Daley, YoungHo Kwon, Hengyao Niu, and Patrick Sung. Focus: 50 years of dna repair: the yale symposium reports: investigations of homologous recombination pathways and their regulation. *The Yale journal of biology and medicine*, 86(4):453–61, 2013.

[29] Stephen Van Komen, Galina Petukhova, Stefan Sigurdsson, Sabrina Stratton, and Patrick Sung. Superhelicity-driven homologous dna pairing by yeast recombination factors rad51 and rad54. *Molecular Cell*, 6(3):563 – 572, 2000.

[30] Brian Munsky, Brooke Trinh, and Mustafa Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(1):318, 2009.

[31] Desmond J. Higham. Modeling and simulating chemical reactions. *SIAM Review*, 50(2):347–368, 2008.

[32] Dirk Fey, Maciej Dobrzynski, and Boris N. Kholodenko. Modeling with ordinary differential equations. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 3. MIT Press, Cambridge, Massachusetts, 2018.

[33] Zoltan Szallasi, Jörg Stelling, and Vipul Periwal. *Modeling Molecular Interaction Networks with Nonlinear Ordinary Differential Equations*, pages 97–123. The MIT Press, 2006.

[34] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem*, 274(42):30169–30181, Oct 1999.

[35] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. *Systems Biology*. Wiley, 2013.

[36] B.P. Ingalls. *Mathematical Modeling in Systems Biology: An Introduction*. MIT Press, 2013.

[37] J. Keener and J. Sneyd. *Mathematical Physiology*, volume 8 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2001.

[38] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[39] Roberto Bertolusso and Marek Kimmel. Kinetic monte carlo analyses of discrete biomolecular events. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 7, pages 113–136. MIT Press, Cambridge, Massachusetts, 2018.

[40] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Chem Phys*, 104:1876–1889, 2000.

[41] A Slepoy, A P Thompson, and S J Plimpton. A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. *J Chem Phys*, 128(20):205101, May 2008.

[42] D T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[43] D T Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[44] D T Gillespie and L Petzold. Improved leap-size selection for accelerated stochastic simulation. *The Journal of Chemical Physics*, 119(16):8229–8234, 2003.

[45] Y Cao, D T Gillespie, and L Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, 124(4):044109, 2006.

[46] Y Cao, D T Gillespie, and L Petzold. Adaptive explicit-implicit tau-leaping method with automatic tau selection. *The Journal of Chemical Physics*, 126(22):224101, 2007.

[47] Y Cao, D. T. Gillespie, and L Petzold. The Slow-Scale Stochastic Simulation Algorithm. *The Journal of Chemical Physics*, 122(014116), January 2005.

[48] C Rao and A P Arkin. Stochastic Chemical Kinetics and the Quasi-Steady-State Assumption: Application to the Gillespie Algorithm. *The Journal of Chemical Physics*, 118:4999–5010, 2003.

[49] L Harris and P Clancy. A "partitioned leaping" approach for multiscale modeling of chemical reaction dynamics. *J. Chem. Phys.*, 125:144107, 2006.

[50] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[51] Brian Munsky. Modeling cellular variability. *Quantitative Biology: From Molecular to Cellular Systems*, page 233, 2012.

[52] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4):044104, 2006.

[53] Brian Munsky and Mustafa Khammash. Transient Analysis of Stochastic Switches and Trajectories with Applications to Gene Regulatory Networks. *IET Systems Biology*, 2(5):323–333, 2008.

[54] Brian Munsky and Mustafa Khammash. The Finite State Projection Approach for the Analysis of Stochastic Noise in Gene Networks. *IEEE Trans. Automat. Contr./IEEE Trans. Circuits and Systems: Part 1*, 52(1):201–214, Jan 2008.

[55] Zachary Fox, Gregor Neuert, and Brian Munsky. Finite state projection based bounds to compare chemical master equation models using single-cell data. *The Journal of Chemical Physics*, 145(7):074101, Aug 2016.

[56] Brian Munsky, Gregor Neuert, and Alexander van Oudenaarden. Using Gene Expression Noise to Understand Gene Regulation. *Science*, 336(6078):183–187, April 2012.

[57] Brian Munsky, Zachary Fox, and Gregor Neuert. Integrating Single-Molecule Experiments and Discrete Stochastic Models to Understand Heterogeneous Gene Transcription Dynamics. *Methods*, 85:12–21, September 2015.

[58] Zachary Fox and Brian Munsky. Stochasticity or noise in biochemical reactions. In Brian Munsky, William S Hlavacek, and Lev S Tsimring, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 5, pages 71–93. MIT Press, Cambridge, Massachusetts, 2018.

[59] K Burrage, M Hegland, S Macnamara, and R Sidje. A Krylov-Based Finite State Projection Algorithm for Solving the Chemical Master Equation Arising in the Discrete Modelling of Biological Systems. *Proc. of The A.A.Markov 150th Anniversary Meeting*, pages 21–37, 2006.

[60] Donald A McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4(3):413–478, 1967.

[61] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[62] C. Vacar, J. Giovannelli, and Y. Berthoumieu. Langevin and hessian with fisher approximation stochastic sampling for parameter estimation of structured covariance. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3964–3967, May 2011.

[63] Jean Peccoud and Bernard Ycart. Markovian modeling of gene-product synthesis. *Theoretical population biology*, 48(2):222–234, 1995.

[64] Leor S Weinberger, John C Burnett, Jared E Toettcher, Adam P Arkin, and David V Schaffer. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 tat fluctuations drive phenotypic diversity. *Cell*, 122(2):169–182, 2005.

[65] Nicolas Le Novere. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics*, 16(3):146–158, 2015.

[66] Yuting Zheng and Ganesh Sriram. Mathematical modeling: bridging the gap between concept and realization in synthetic biology. *Journal of Biomedicine and Biotechnology*, 2010, 2010.

[67] Sara Hamis, Gibin G Powathil, and Mark AJ Chaplain. Blackboard to bedside: a mathematical modeling bottom-up approach toward personalized cancer treatments. *JCO clinical cancer informatics*, 3:1–11, 2019.

[68] Uri Goldsztejn, David Schwartzman, and Arye Nehorai. Public policy and economic dynamics of COVID-19 spread: A mathematical modeling study. *PloS one*, 15(12):e0244174, 2020.

[69] Pierre-François Verhulst. Notice sur la loi que la population suit dans son accroissement. *Correspondence mathematique et physique*, 10:113–129, 1838.

[70] CM Guldberg and P Waage. Avhandl. *Norske Videnskaps-Akademi Oslo, Mat. Naturv. Kl*, page 35, 1864.

[71] Leonor Michaelis, Maud L Menten, et al. Die kinetik der invertinwirkung. *Biochem. z*, 49(333-369):352, 1913.

[72] Alfred J Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14(3):271–274, 1910.

[73] Vito Volterra. Variations and fluctuations of the number of individuals in animal species living together. *Animal ecology*, pages 412–433, 1931.

[74] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.

[75] Derek A Roff. *Evolutionary quantitative genetics*. Springer Science & Business Media, 2012.

[76] Izet Masic. The most influential scientists in the development of medical informatics (13): Margaret belle dayhoff. *Acta Informatica Medica*, 24:299, 2016.

[77] Jeff Gauthier, Antony T Vincent, Steve J Charette, and Nicolas Derome. A brief history of bioinformatics. *Briefings in bioinformatics*, 20(6):1981–1996, 2019.

[78] Alan S. Perelson, Avidan U. Neumann, Martin Markowitz, John M. Leonard, and David D. Ho. HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271(5255):1582–1586, 1996.

[79] Karolin Luger, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature*, 389(6648):251–260, 1997.

[80] Edda Klipp and Wolfram Liebermeister. Mathematical modeling of intracellular signaling pathways. *BMC neuroscience*, 7:1–16, 2006.

[81] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[82] Jonathan R Karr, Jayodita C Sanghvi, Derek N Macklin, Miriam V Gutschow, Jared M Jacobs, Benjamin Bolival, Nacyra Assad-Garcia, John I Glass, and Markus W Covert. A

whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[83] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[84] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[85] Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, et al. Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873):590–596, 2021.

[86] Jennifer A. Doudna and Emmanuelle Charpentier. The new frontier of genome engineering with CRISPR-cas9. *Science*, 346(6213), 2014.

[87] Jeffry D. Sander and J. Keith Joung. CRISPR-cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 32:347–355, 2014.

[88] Tianhai Tian, Songlin Xu, Junbin Gao, and Kevin Burrage. Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics*, 23(1):84–91, 10 2006.

[89] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[90] Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 07 2001.

[91] Daniel T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 07 2000.

[92] Lisa Weber, William Raymond, and Brian Munsky. Identification of gene regulation models from single-cell data. *Physical Biology*, 15(5):055001, May 2018.

[93] Lisa Weber, William Raymond, and Brian Munsky. Tutorial on the identification of gene regulation models from single-cell data. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 30. MIT Press, Cambridge, Massachusetts, 2018.

[94] Brian Munsky and Gregor Neuert. From analog to digital models of gene regulation. *Physical Biology*, 12(4):045004, July 2015.

[95] Brian Munsky, William S Hlavacek, and Lev S Tsimring. *Quantitative Biology: Theory, Computational Methods, and Models*. MIT Press, 2018.

[96] N.G. VanKampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 3rd ed. edition, 2007.

[97] Khanh N Dinh and Roger B Sidje. Understanding the finite state projection and related methods for solving the chemical master equation. *Physical Biology*, 13(3):035003, may 2016.

[98] Karol Nienaltowski, Tomasz Jetka, and Michal Komorowski. Sensitivity analysis. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 14. MIT Press, Cambridge, Massachusetts, 2018.

[99] Tomasz Jetka, Karol Nienaltowski, and Michal Komorowski. Experimental design. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 15. MIT Press, Cambridge, Massachusetts, 2018.

[100] Bryan C. Daniels, Maciej Dobrzynski, and Dirk Fey. Parameter estimation, sloppiness, and model identifiability. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Modelss*, chapter 13. MIT Press, Cambridge, Massachusetts, 2018.

[101] Zach Fox and Brian Munsky. Bayesian parameter estimation and markov chain monte carlo. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 16. MIT Press, Cambridge, Massachusetts, 2018.

[102] Bence Mélykúti, Elias August, Antonis Papachristodoulou, and Hana El-Samad. Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Systems Biology*, 4(38), 2010.

[103] Joshua F Apgar, Jared E Toettcher, Drew Endy, Forest M White, and Bruce Tidor. Stimulus design for model selection and validation in cell signaling. *PLOS Computational Biology*, 4(2):1–10, 02 2008.

[104] J Peccoud and B Ycart. Markovian Modeling of Gene-Product Synthesis. *Theoretical Population Biology*, 48(2):222–234, October 1995.

[105] Dale Muzzey, Carlos A. Gómez-Uribe, Jerome T. Mettetal, and Alexander van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138:160–171, July 2009.

[106] Jerome T. Mettetal, Dale Muzzey, Carlos Gómez-Uribe, and Alexander van Oudenaarden. The frequency dependence of osmo-adaptation in <i>saccharomyces cerevisiae</i>. *Science*, 319(5862):482–484, 2008.

[107] Guoliang Li, Benjamin K. Kesler, Alexander Thiemicke, Dustin C. Rogers, and Gregor Neuert. Linearly changing stress environment causes cellular growth phenotype. *bioRxiv*, 2017.

[108] Martin Krzywinski and Naomi Altman. Points of significance: importance of being uncertain, 2013.

[109] JA Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308, 1965.

[110] G. B Dantzig. *Linear Programming and Extensions*. Princeton, NJ: Princeton University Press, 1963.

[111] Peter JM Van Laarhoven and Emile HL Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.

[112] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10):1871–78, October 2007.

[113] Bryan C Daniels, Yan-Jiun Chen, James P Sethna, Ryan N Gutenkunst, and Christopher R Myers. Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, 19(4):389–395, 2008.

[114] Margaritis Voliotis, Philipp Thomas, Clive G. Bowsher, and Ramon Grima. The extra reaction algorithm for stochastic simulation of biochemical reaction systems in fluctuating environments. In Brian Munsky, Lev S. Tsimring, and William S. Hlavacek, editors, *Quantitative Biology: Theory, Computational Methods and Examples of Models*, chapter 8. MIT Press, Cambridge, Massachusetts, 2018.

[115] Margaritis Voliotis, Philipp Thomas, Ramon Grima, and Clive G Bowsher. Stochastic simulation of biomolecular networks in dynamic environments. *PLoS Computational Biology*, 12(6):e1004923, 2016.

[116] Juliane Liepe, Sarah Filippi, Michał Komorowski, and Michael P. H. Stumpf. Maximizing the information content of experiments in systems biology. *PLOS Computational Biology*, 9(1):1–13, 01 2013.

[117] Jakob Ruess, Andreas Milias-Argeitis, and John Lygeros. Designing experiments to understand the variability in biochemical reaction networks. *Journal of The Royal Society Interface*, 10(88):20130588, 2013.

[118] Brian Munsky, Brooke Trinh, and Mustafa Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(1):318, 2009.

[119] Chunbo Lou, Brynne Stanton, Ying-Ja Chen, Brian Munsky, and Christopher A Voigt. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature Biotechnology*, 30:1137–1142, 2012.

[120] R.P. Bowater, F. Aboul-ela, and D.M. Lilley. Large-scale opening of a + t rich regions within supercoiled dna molecules is suppressed by salt. *Nucleic Acids Research*, 32:2042–50, 1994.

[121] A. Lal, A. Dhar, A. Trostel, F. Kouzine, A.S.N. Seshasayee, and S. Adhya. Genome scale patterns of supercoiling in a bacterial chromosome. *Nat. Commun.*, 7:11055, 2016.

[122] F. Kouzine, A. Gupta, L. Baranello, D. Wojtowicz, K. Ben-Aissa, J. Liu, T.M. Przytycka, and D. Levens. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.*, 20:396–403, 2013.

[123] C. Naughton, N. Avlonitis, S. Corless, J.G. Prendergast, I.K. Mati, P.P. Eijk, S.L. Cockroft, N. Bradley, B. Ylstra, and N. Gilbert. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.*, 20:387–95, 2013.

[124] S.S. Teves and S. Henikoff. Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.*, 21:88–94, 2014.

[125] Hong Liu, Niveen Mulholland, Haiqing Fu, and Keji Zhao. Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling. *Molecular and Cellular Biology*, 26(7):2550–2559, 2006.

[126] J. Zhou, S.K. Gregurick, S. Krueger, and F.P. Schwarz. Conformational changes in single-strand dna as a function of temperature by SANS. *Biophysical Journal*, 90(2):544 – 551, 2006.

[127] Mika A. Kastenholz, Thomas U. Schwartz, and Philippe H. Hünenberger. The transition between the b and z conformations of DNA investigated by targeted molecular dynamics simulations with explicit solvation. *Biophysical Journal*, 91(8):2976 – 2990, 2006.

[128] Sangeeta Kundu, Sanchita Mukherjee, and Dhananjay Bhattacharyya. Effect of temperature on DNA double helix: An insight from molecular dynamics simulation. *Journal of Biosciences*, 37(3):445–455, Jul 2012.

[129] Korbinian Liebl and Martin Zacharias. Unwinding induced melting of double-stranded DNA studied by free energy simulations. *The Journal of Physical Chemistry B*, 121(49):11019–11030, 2017. PMID: 29064703.

[130] Monrudee Liangruksa and Somchai Wongwises. An elastic model of DNA under thermal induced stress. *Mathematical Biosciences*, 300:47 – 54, 2018.

[131] Douglas Poland. DNA probability profiles: examples from the treponema pallidum genome. *Biophysical Chemistry*, 104(1):279 – 289, 2003.

[132] Ping Xie. A model of DNA unwinding dynamics by the RecBCD complex and its regulation by chi recognition. *Journal of Theoretical Biology*, 448:142 – 156, 2018.

[133] Dahai Gai, Y Paul Chang, and Xiaojiang S Chen. Origin DNA melting and unwinding in DNA replication. *Current Opinion in Structural Biology*, 20(6):756 – 762, 2010.

[134] Soumadwip Ghosh, Himanshu Dixit, and Rajarshi Chakrabarti. Ion assisted structural collapse of a single stranded DNA: A molecular dynamics approach. *Chemical Physics*, 459:137 – 147, 2015.

[135] Tamar Schlick and Wilma K. Olson. Supercoiled DNA energetics and dynamics by computer simulation. *Journal of Molecular Biology*, 223(4):1089 – 1119, 1992.

[136] Eric J. Sorin, Young Min Rhee, Bradley J. Nakatani, and Vijay S. Pande. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophysical Journal*, 85(2):790 – 803, 2003.

[137] Srinivasaraghavan Kannan, Kai Kohlhoff, and Martin Zacharias. B-DNA under stress: Over- and untwisting of DNA during molecular dynamics simulations. *Biophysical Journal*, 91(8):2956 – 2965, 2006.

[138] M. Kruithof and J. van Noort. Hidden markov analysis of nucleosome unwrapping under force. *Biophysical Journal*, 96(9):3708 – 3715, 2009.

[139] Yunqiang Bian, Feng Song, Zanxia Cao, Liling Zhao, Jiafeng Yu, Xinlu Guo, and Jihua Wang. Fast-folding pathways of the thrombin-binding aptamer g-quadruplex revealed by a markov state model. *Biophysical Journal*, 114(7):1529 – 1538, 2018.

[140] Marco Zoli. Twist versus nonlinear stacking in short DNA molecules. *Journal of Theoretical Biology*, 354:95 – 104, 2014.

[141] John F. Beausang, Chiara Zurla, Carlo Manzo, David Dunlap, Laura Finzi, and Philip C. Nelson. DNA looping kinetics analyzed using diffusive hidden markov model. *Biophysical Journal*, 92(8):L64 – L66, 2007.

[142] Daniel P. Aalberts and Nathan O. Hodas. Efficient computation of optimal oligo–RNA binding. *Nucleic Acids Research*, 32(22):6636–6642, 01 2004.

[143] Jin Yu, Wei Cheng, Carlos Bustamante, and George Oster. Coupling translocation with nucleic acid unwinding by NS3 helicase. *Journal of Molecular Biology*, 404(3):439 – 455, 2010.

[144] Mrinal Kanti Si, Sumit Kumar Pramanik, and Bishwajit Ganguly. Tuning the ring strain effect in acridine derivatives on binding affinity with g-quadruplex-DNA: A computational and experimental study. *International Journal of Biological Macromolecules*, 124:1177 – 1185, 2019.

[145] Federico De Masi, Christian A. Grove, Anastasia Vedenko, Andreu Alibés, Stephen S. Gisselbrecht, Luis Serrano, Martha L. Bulyk, and Albertha J. M. Walhout. Using a structural and logics systems approach to infer bHLH–DNA binding specificity determinants. *Nucleic Acids Research*, 39(11):4553–4563, 02 2011.

[146] Chaitanya Rastogi, H. Tomas Rube, Judith F. Kribelbauer, Justin Crocker, Ryan E. Loker, Gabriella D. Martini, Oleg Laptenko, William A. Freed-Pastor, Carol Prives, David L. Stern, Richard S. Mann, and Harmen J. Bussemaker. Accurate and sensitive quantification of protein-DNA binding affinity. *Proceedings of the National Academy of Sciences*, 115(16):E3692–E3701, 2018.

[147] Justin B. Kinney, Gašper Tkačik, and Curtis G. Callan. Precise physical models of protein–DNA interaction from high-throughput data. *Proceedings of the National Academy of Sciences*, 104(2):501–506, 2007.

[148] Jason D Buenrostro, Carlos L Araya, Lauren M Chircus, Curtis J Layton, Howard Y Chang, Michael P Snyder, and William J Greenleaf. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nature biotechnology*, 32(6):562, 2014.

[149] Giulia Palermo, Lorenzo Casalino, Alessandra Magistrato, and J. Andrew McCammon. Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *Journal of Structural Biology*, 2019.

[150] Nawid Zarrabi, Peter Schluesche, Michael Meisterernst, Michael Börsch, and Don C. Lamb. Analyzing the dynamics of single TBP-DNA-NC2 complexes using hidden markov models. *Biophysical Journal*, 115(12):2310 – 2326, 2018.

[151] G. A. Held, G. Grinstein, and Y. Tu. Modeling of DNA microarray data by using physical properties of hybridization. *Proceedings of the National Academy of Sciences*, 100(13):7575–7580, 2003.

[152] Haris Vikalo and Mahsuni Gokdemir. An MCMC algorithm for target estimation in real-time DNA microarrays. *EURASIP Journal on Advances in Signal Processing*, 2010(1):736301, Aug 2010.

[153] M. Gokdemir and H. Vikalo. A particle filtering algorithm for parameter estimation in real-time biosensor arrays. In *2009 IEEE International Workshop on Genomic Signal Processing and Statistics*, pages 1–4, May 2009.

[154] T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, and M. Girolami. Langevin diffusions and the metropolis-adjusted langevin algorithm. *Statistics & Probability Letters*, 91:14 – 19, 2014.

[155] Jonas Wallin and David Bolin. Efficient adaptive MCMC through precision estimation. *Journal of Computational and Graphical Statistics*, 27(4):887–897, 2018.

[156] A. van den Bos. *Parameter Estimation for Scientists and Engineers*. Wiley, 2007.

[157] John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *IN BAYESIAN STATISTICS*, pages 169–193. University Press, 1992.

[158] D Kowalski, D A Natale, and M J Eddy. Stable DNA unwinding, not "breathing," accounts for single-strand-specific nuclease hypersensitivity of specific a+t-rich sequences. *Proceedings of the National Academy of Sciences*, 85(24):9464–9468, 1988.

[159] Dina Zhabinskaya and Craig J. Benham. Theoretical analysis of competing conformational transitions in superhelical DNA. *PLOS Computational Biology*, 8(4):1–21, 04 2012.

[160] Juhong Liu, Fedor Kouzine, Zuqin Nie, Hye-Jung Chung, Zichrini Elisha-Feil, Achim Weber, Keji Zhao, and David Levens. The FUSE/FBP/FIR/TFIIH system is a molecular machine programming a pulse of c-myc expression. *The EMBO Journal*, 25(10):2119–2130, 2006.

[161] Silvana Di Palma, Maryou B K Lambros, Kay Savage, Chris Jones, Alan Mackay, Tim Dexter, Marjan Iravani, Kerry Fenwick, Alan Ashworth, and Jorge S Reis-Filho. Oncocytic change in pleomorphic adenoma: molecular evidence in support of an origin in neoplastic cells. *Journal of Clinical Pathology*, 60(5):492–499, 2007.

[162] Lowell G. Sheflin and David Kowalski. Altered DNA conformations detected by mung bean nuclease occur in promoter and terminator regions of supercoiled pBR322 DNA. *Nucleic Acids Research*, 13(17):6137–6154, 09 1985.

[163] A.J.M. Matzke, T.M. Weiger, and M. Matzke. Ion channels at the nucleus: Electrophysiology meets the genome. *Mol. Plant*, 3(4):642–52, 2010.

[164] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

[165] Venkat Gudivada, Amy Apon, and Junhua Ding. Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, 10:1–20, 07 2017.

[166] Shayan Doroudi. The bias-variance tradeoff: How data science can inform educational debates. *AERA Open*, 6(4):2332858420977208, 2020.

[167] Caterina Thomaseth and Nicole E. Radde. The effect of normalisation and error model choice on the distribution of the maximum likelihood estimator for a biochemical reaction. *IET Systems Biology*, 17(1):1–13, 2023.

[168] Lisa M. Weber and Rebecca A. Atadero. Work in progress: Incorporation of diversity and inclusion into the undergraduate chemical engineering curriculum. In *2021 Collaborative Network for Engineering and Computing Diversity (CoNECD) Conference*, Virtual - 1pm to 5pm Eastern Time Each Day, January 2021. ASEE Conferences. https://peer.asee.org/36145.

[169] Equity, inclusion and diversity in engineering: Why they matter. https://www.colorado.edu/emp/2021/01/22/equity-inclusion-and-diversity-engineering-why-they-matter. Accessed on 20 Nov 2023.

[170] William W. Wulf. Diversity in engineering. *Leadership and Management in Engineering*, 1(4):31–35, 2001.

[171] David C. Mays. How diversity makes better engineering teams. *Journal AWWA*, 114(7):62–67, 2022.

[172] Carroll Seron, Susan Silbey, Erin Cech, and Brian Rubineau. “i am not a feminist, but. . .”: Hegemony of a meritocratic ideology and the limits of critique among women in engineering. *Work and Occupations*, 45(2):131–167, 2018.

[173] Cynthia Rice and David Mays. Opinion: Building diversity, equity, and inclusion into an engineering course. *Advances in Enginering Education*, 10(4), 2022.

[174] Stephanie Farrell, Allison Godwin, and Donna M Riley. A sociocultural learning framework for inclusive pedagogy in engineering. *Chemical Engineering Education*, 55(4):192–204, 2021.

[175] Erin A. Cech. Culture of disengagement in engineering education? *Science, Technology, & Human Values*, 39(1):42–72, 2014.

[176] U.S. Census Bureau. Detailed occupation for the civilian employed male population 16 years and over. https://data.census.gov/table/ACSDT1Y2022.B24115?q=b24115. Accessed on 5 November 2023.

[177] U.S. Census Bureau. Detailed occupation for the civilian employed female population 16 years and over. https://data.census.gov/table/ACSDT1Y2022.B24116?q=b24116. Accessed on 5 November 2023.

[178] US DOE. Digest of Education Statistics Table 505.30. Technical report, Integrated Postsecondary Education Data System (IPEDS), Washingnton, D.C., 2021.

[179] Mary Frank Fox, Gerhard Sonnert, and Irina Nikiforova. Successful programs for undergraduate women in science and engineering: Adapting versus adopting the institutional environment. *Research in Higher Education*, 50:333–353, June 2009.

[180] Lauren Drankoff, Sandra L. Furterer, and Elizabeth Hart. Undergraduate women in science and engineering mentoring program to enhance gender diversity demonstrates success during the COVID pandemic. In *2021 ASEE Virtual Annual Conference Content Access*, Virtual Conference, July 2021. ASEEASEE Conferences. https://peer.asee.org/37952.

[181] Suzanne M. Kresta. Teaching innovation in an age of disruption. *The Canadian Journal of Chemical Engineering*, 99(10):2138–2148, 2021.

[182] Stephen Secules and Stephanie Masta. Towards a framework for equity in engineering classrooms. In *2020 IEEE Frontiers in Education Conference (FIE)*, pages 1–4, 2020.

[183] Christine E. Sleeter. Confronting the marginalization of culturally responsive pedagogy. *Urban Education*, 47(3):562–584, 2012.

[184] Ilene J Busch-Vishniac and Jeffrey P. Jarosz. Can diversity in the undergraduate engineering population be enhanced through curricular change. *Journal of Women and Minorities in Science and Engineering*, 10:255–282, 2004.

[185] Sonia Bansal, Aaron M. Kyle, Andrew O. Brightman, and Jennifer R. Amos. Approaches to address new ABET diversity, equity, and inclusion criteria in biomedical engineering curricula. *Biomedical Engineering Education*, 3:331–344, July 2023.

[186] Dustin B. Thoman, Melo-Jean Yap, Felisha A. Herrera, and Jessi L. Smith. Diversity interventions in the classroom: From resistance to action. *CBE—Life Sciences Education*, 20(4):ar52, 2021. PMID: 34546104.

[187] Arianne Collopy, Heather Johnson, Katherine Goodman, Tom Altman, Maryam Darbeheshti, Kristin Wood, and David Mays. Exploring nudging approaches for growing a culture of diversity and inclusion with engineering faculty. In *2022 ASEE Annual Conference & Exposition*, Minneapolis, MN, August 2022. ASEE Conferences. https://strategy.asee.org/42018.

[188] Rion J. Wendland and Kristan S. Worthington. Improv activities in an engineering classroom increase student self-perceptions of engagement, adaptability, and communication. *Biomedical Engineering Education*, November 2023.

[189] Karen E. Rambo-Hernandez, Abhik Roy, Melissa Lynn Morris, Robin A. M. Hensel, Jeremy Clinton Schwartz, Rebecca A. Atadero, and Christina Paguyo. Using interactive theater to promote inclusive behaviors in teams for first-year engineering students: A sustainable approach. In *2018 CoNECD - The Collaborative Network for Engineering and Computing Diversity Conference*, Crystal City, Virginia, April 2018. ASEE Conferences. https://peer.asee.org/29592.

[190] Susie Huggins and Robin A.M. Hensel. Engendering inclusion by implementing an interactive theatre sketch. In *2023 ASEE North Central Section Conference*, Morgantown, West Virginia, March 2023. ASEE Conferences. https://peer.asee.org/44691.

[191] Karen E. Rambo-Hernandez, Melissa Lynn Morris, Anne Marie Aramati Casper, Robin A. M. Hensel, Jeremy Clinton Schwartz, and Rebecca A. Atadero. Examining the ef-

fects of equity, inclusion, and diversity activities in first-year engineering classes. In *2019 ASEE Annual Conference & Exposition*, Tampa, Florida, June 2019. ASEE Conferences. https://peer.asee.org/32782.

[192] Nilanjana Dasgupta, Melissa McManus Scircle, and Matthew Hunsinger. Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, 112(16):4988–4993, 2015.

[193] Akiko Nakamura. Fostering diversity and inclusion and understanding implicit bias in undergraduate chemical education. *Journal of Chemical Education*, 99(1):331–337, 2022.

[194] Jake Nease, Vincent Leung, Shelir Ebrahimi, Beth Levinson, Ishwar K. Puri, and Carlos D. M. Filipe. A learner's journey towards a chemical engineering degree. *The Canadian Journal of Chemical Engineering*, 99(10):2149–2162, 2021.

[195] Brandon Reynante. Learning to design for social justice in community-engaged engineering. *Journal of Engineering Education*, 111(2):338–356, 2022.

[196] Colorado State University Office of Institutional Research, Planning, and Effectiveness. Enrollment demographics for engineering. https://www.ir.colostate.edu/data-reports/students/. Accessed on 6 November 2023.

[197] E. A. Cech and T. J. Waidzunas. Systemic inequalities for lgbtq professionals in stem. *Science Advances*, 7(3):eabe0933, 2021.

[198] Evan: MATLAB Central File Exchange. Plot groups of stacked bars. https://www.mathworks.com/matlabcentral/fileexchange/32884-plot-groups-of-stacked-bars. Retrieved 09 November 2023.

[199] Csu mechanical engineering department initiatives. https://www.engr.colostate.edu/me/diversity-equity-and-inclusion/. Accessed on 19 November 2023.

[200] University of colorado at boulder demographics. https://public.tableau.com/app/profile/university.of.colorado.boulder.ir/vizzes. Accessed: 3 December 2023.

[201] State Demography Office. Single year of age by race / ethnicity: Colorado, 2022. https://demography.dola.colorado.gov/assets/html/racebyage.html#. Accessed on 19 November 2023.

[202] Cheryl Staats, Kelly Capatosto, Robin A. Wright, and Victoria W. Jackson. State of the science: Implicit bias review. Technical report, The Ohio State University: Kirwan Institute for the Study of Race and Ethnicity, 2016.

[203] Elaine Seat, J. Roger Parsons, and William A. Poppen. Enabling engineering performance skills: A program to teach communication, leadership, and teamwork*. *Journal of Engineering Education*, 90(1):7–12, 2001.

[204] Advisory Committee to the National Science Foundation. Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology. Technical Report NSF 96-139, National Science Foundation (NSF), 1996.

[205] Carolyn Meyers. Restructuring engineering education: A focus on change. Report of an NSF workshop on Engineering Education NSF 95-65, National Science Foundation (NSF), April 1995.

[206] ABET Engineering Accreditation Commission (EAC). Comparison of proposal submitted in 2015 to proposal submitted in 2016. Technical report, ABET, 2016.

[207] Carroll Seron, Susan S. Silbey, Erin Cech, and Brian Rubineau. Persistence is cultural: Professional socialization and the reproduction of sex segregation. *Work and Occupations*, 43(2):178–214, 2016.

[208] Franziska Meissner, Laura Anne Grigutsch, Nicolas Koranyi, Florian Müller, and Klaus Rothermund. Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions. *Frontiers in Psychology*, 10, 2019.

[209] Benedek Kurdi, Allison E Seitchik, Jordan R Axt, Timothy J Carroll, Arpi Karapetyan, Neela Kaushik, Diana Tomezsko, Anthony G Greenwald, and Mahzarin R Banaji. Relationship between the implicit association test and intergroup behavior: A meta-analysis. *American psychologist*, 74(5):569, 2019.

[210] Bertram Gawronski and Adam Hahn. Implicit measures: Procedures, use, and interpretation. In *Measurement in social psychology*, pages 29–55. Routledge, 2018.

[211] Bertram Gawronski, Robert J. Rydell, Jan De Houwer, Skylar M. Brannon, Yang Ye, Bram Vervliet, and Xiaoqing Hu. Chapter one - contextualized attitude change. In James M. Olson, editor, *Advances in Experimental Social Psychology*, volume 57 of *Advances in Experimental Social Psychology*, pages 1–52. Academic Press, 2018.

[212] Sydney Alderfer, Rachel McMillan, Katlyn Murphy, and Nicole Kelp. Inclusive science communication training for first-year STEM students promotes their identity and self-efficacy as scientists and science communicators. *Frontiers in Education*, 8, 2023.

[213] Randy Vickery, Katlyn Murphy, Rachel McMillan, Sydney Alderfer, Jasmine Donkoh, and Nicole Kelp. Analysis of inclusivity of published science communication curricula for scientists and STEM students. *CBE—Life Sciences Education*, 22(1):ar8, 2023. PMID: 36637377.

[214] Katlyn M. Murphy and Nicole C. Kelp. Undergraduate STEM students' science communication skills, science identity, and science self-efficacy influence their motivations and behaviors in STEM community engagement. *Journal of Microbiology and Biology Education*, 24(1):e00182–22, 2023.

[215] Bettina J. Casad, Danielle L. Oyler, Erin T. Sullivan, Erika M. McClellan, Destiny N. Tierney, Drake A. Anderson, Paul A. Greeley, Michael A. Fague, and Brian J. Flammang. Wise psychological interventions to improve gender and racial equality in STEM. *Group Processes and Intergroup Relations*, 21(5):767–787, 2018.

[216] Yoi Tibbetts, Judith M. Harackiewicz, Stacy J. Priniski, and Elizabeth A. Canning. Broadening participation in the life sciences with social–psychological interventions. *CBE—Life Sciences Education*, 15(3):es4, 2016. PMID: 27543632.

[217] Gloria Ladson-Billings. Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3):465–491, 1995.

[218] Brit Shields. Justice, equity, diversity, and inclusion curriculum within an introductory bioengineering course. *Biomedical Engineering Education*, 3(1):39–49, 2023.

[219] Barbara J. Grosz, David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. Embedded ethics: Integrating ethics across cs education. *Commun. ACM*, 62(8):54–61, jul 2019.

[220] Carl Mitcham and Elaine Englehardt. Ethics across the curriculum: Prospects for broader (and deeper) teaching and learning in research and engineering ethics. *Science and engineering ethics*, 25, 12 2019.

[221] Anthony Butterfield, Alon McCormick, and Stephanie Farrell. Building LGBTQ-inclusive chemical engineering classrooms and departments. *Chemical Engineering Education*, 52(2):107–113, 2018.

[222] David S. Yeager and Gregory M. Walton. Social-psychological interventions in education: They're not magic. *Review of Educational Research*, 81(2):267–301, 2011.

[223] Mary M. Staehle, Rachel C. Childers, and William H. Guilford. From theory to practice: Unraveling the impact of experiential learning in biomedical engineering education. *Biomedical Engineering Education*, 3:111–112, July 2023.

[224] William H. Guilford, Miiri Kotche, and Rachael H. Schmedlen. A survey of clinical immersion experiences in biomedical engineering. *Biomedical Engineering Education*, 3:113–122, July 2023.

[225] Justin Huber, Steven Higbee, Christina Espinosa, Babak Bazrgari, and Sharon Miller. Immersion experiences for biomedical engineering undergraduates: Comparing strategies and local partnerships at two institutions. *International Journal of Engineering Education*, 39:961–975, 2023.

[226] Sudeshna Pal, Mark Steiner, Alain J. Kassab, Patrick Pabian, Adam Golden, and Edward A. Ross. An experiential learning based design program in rehabilitation engineering. *Biomedical Engineering Education*, 3:193–199, 2023.

[227] Luis Montesinos, David Ernesto Salinas-Navarro, and Alejandro Santos-Diaz. Transdisciplinary experiential learning in biomedical engineering education for healthcare systems improvement. *BMC Medical Education*, 23:207, 2023.

[228] Luis Montesinos, Alejandro Santos-Diaz, David E. Salinas-Navarro, and Leopoldo Cendejas-Zaragoza. Experiential learning in biomedical engineering education using wearable devices: A case study in a biomedical signals and systems analysis course. *Education Sciences*, 12(9), 2022.

[229] Neil Williams, Humma Hussain, Pinali Manojkumar, and Asim Thapa. An evaluation of a STEM summer undergraduate research internship scheme : student-perceived learning gains. *New directions in the teaching of physical sciences*, 11, 2016.

[230] Stephanie M. George and Zachary J. Domire. A Six-Year Review of the Biomedical Engineering in Simulations, Imaging, and Modeling Undergraduate Research Experience. *Journal of Biomechanical Engineering*, 142(11):111012, 09 2020.

[231] A. F. Chou, D. Hammon, and D. R. Akins. Impact and outcomes of the oklahoma idea network of biomedical research excellence summer undergraduate research program. *Journal of microbiology and biology education*, 20, 2019.

[232] Annie Wang Cassandra Sue Ellen Jamison, Jacob Fuher and Aileen Huang-Saad. Experiential learning implementation in undergraduate engineering education: a systematic search and review. *European Journal of Engineering Education*, 47(6):1356–1379, 2022.

[233] Megan Huffstickler, Sarah E. Zappe, Keefe B. Manning, and Margaret J. Slattery. Impact of a biomedical engineering undergraduate research program on student and faculty perceptions of creativity. In *2017 ASEE Annual Conference and Exposition*, Columbus, Ohio, June 2017. ASEE Conferences. https://peer.asee.org/28456.

[234] Jean Lave and Etienne Wenger. Situated learning: Legitimate peripheral participation. In *Learning in Doing: Social, Cognitive and Computational Perspectives*. Cambridge University Press, Cambridge: Cambridge University Press, 1991.

[235] L. S. Vygotsky. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press, 1978.

[236] Barry J Wadsworth. *Piaget's theory of cognitive and affective development: Foundations of constructivism.* Longman Publishing, 1996.

[237] Melissa Lynn Morris, Joseph Dygert, and Robin A.M. Hensel. How do student perceptions of engineers and engineering as a career relate to their self-efficacy, career expectations, and grittiness? In *2020 ASEE Virtual Annual Conference Content Access*, Virtual On line, June 2020. ASEE Conferences. https://peer.asee.org/34729.

[238] Robin Hensel, Joseph Dygert, and Melissa Morris. Academic success and retention of underprepared students. In *2021 ASEE Virtual Annual Conference Content Access*, Virtual Conference, July 2021. ASEE Conferences.

[239] Harriet Hartman, Ralph Alan Dusseau, Beena Sukumaran, Stephanie Farrell, Tiago R. Forin, Stephanie Lezotte, Kauser Jahan, and Sarah K. Bauerand Danilo Zeppilli. First-generation college students and othering in undergraduate engineering. In *2019 ASEE*

Annual Conference and Exposition, Tampa, Florida, June 2019. ASEE Conferences. https://peer.asee.org/32844.

[240] Nicole C. Kelp, Melissa McCartney, Mark A. Sarvary, Justin F. Shaffer, and Michael J. Wolyniak. Developing science literacy in students and society: Theory, research, and practice. *Journal of Microbiology and Biology Education*, 24(2):e00058–23, 2023.

[241] Renee M Desing, Amena S Shermadou, Tanya M Nocera, David A Delaine, S Zahra Atiq, Alexis Ortiz-Rosario, Rachel C Childers, and Racheida S Lewis. Identifying supports and barriers in engineering students' processes toward career attainment. *Biomedical Engineering Education*, pages 1–24, 08 2023.

[242] Caley J. Valdez and Nicole C. Kelp. Student perceptions of inclusive pedagogy in undergraduate STEM classrooms. *Journal of Microbiology and Biology Education*, 0(0):e00097–23, October 2023.

[243] Seoyeon Park, Rebecca A. Atadero, Anne Marie Aramati Casper, Karen E. Rambo-Hernandez, Jody Paul, Melissa Lynn Morris, Christopher Douglas Griffin, Ronald R. DeLyser, Christina Paguyo, and Scott T. Leutenegger. Partnership for equity: Engaging with faculty to cultivate inclusive professional identities for engineers and computer scientists. In *2020 ASEE Virtual Annual Conference Content Access*, Virtual On line, June 2020. ASEE Conferences. https://peer.asee.org/35040.

[244] Harriet Hartman, Tiago Forin, Beena Sukumaran, Stephanie Farrell, Parth Bhavsar, Kauser Jahan, Ralph Dusseau, Theresa Bruckerhoff, Peter Cole, Stephanie Lezotte, Danilo Zeppilli, and Danielle Macey. Strategies for improving diversity and inclusion in an engineering department. *Journal of Professional Issues in Engineering Education and Practice*, 145(2):04018016, 2019.

[245] Stephanie Lezotte. ‘we’re supposed to be at the forefront’: a multiple case study exploring how institutional context shapes engineering diversity and inclusion initiatives. *Engineering Studies*, 15(3):221–239, 2023.

[246] Mary Frank Fox, Gerhard Sonnert, and Irina Nikiforova. Programs for undergraduate women in science and engineering: Issues, problems, and solutions. *Gender and Society*, 25(5):589–615, 2011.

[247] Jennifer A Richeson and Richard J Nussbaum. The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, 40(3):417–423, 2004.

[248] Victoria C. Plaut, Kecia M. Thomas, and Matt J. Goren. Is multiculturalism or color blindness better for minorities? *Psychological Science*, 20(4):444–446, 2009. PMID: 19399972.

[249] Evan P. Apfelbaum, Michael I. Norton, and Samuel R. Sommers. Racial color blindness: Emergence, practice, and implications. *Current Directions in Psychological Science*, 21(3):205–209, 2012.

[250] David Sparkman. *Multiculturalism, Colorblindness, and Prejudice: Examining How Diversity Ideologies Impact Intergroup Attitudes*. PhD thesis, University of Arkansas, Fayetteville, 2018. Retrieved from https://scholarworks.uark.edu/etd/2923.

[251] Anastasia Batkhina, John W. Berry, Tomas Jurcik, Dmitrii Dubrov, and Dmitry Grigoryev. More similarity if different, more difference if similar: Assimilation, colorblindness, multiculturalism, polyculturalism, and generalized and specific negative intergroup bias. *Europe’s Journal of Psychology*, 18(4):369–390, 2022.

[252] Victoria C. Plaut, Kecia M. Thomas, Kyneshawau Hurd, and Celina A. Romano. Do color blindness and multiculturalism remedy or foster discrimination and racism? *Current Directions in Psychological Science*, 27(3):200–206, 2018.

[253] Colorado department of higher education equity toolkit. https://cdhe.colorado.gov/equity-toolkit, 2019. Accessed: 26 November 2023.

[254] Paulo Freire. *Pedagogy of the Oppressed.* New York: Seabury Press, 1968.

[255] Nicole C. Kelp and Breeana Hubbard. Scaffolded curriculum for developing science communication skills in life science undergraduates. *Journal of Microbiology and Biology Education*, 22(1):10.1128/jmbe.v22i1.2255, 2021.

[256] Purdue EPICS program. https://engineering.purdue.edu/EPICS/about. Accessed: 30 November 2023.

[257] GA tech VIP: The vertically integrated projects program. https://www.vip.gatech.edu/vip-vertically-integrated-projects-program. Accessed: 30 November 2023.

[258] VIP consortium: Vertically integrated projects. https://www.vip-consortium.org/content/vip-consortium. Accessed: 30 November 2023.

[259] NYU: Vertically integrated projects. https://engineering.nyu.edu/research-innovation/student-research/vertically-integrated-projects. Accessed: 30 November 2023.

Appendix A

CSU Demographic Statistics for Undergraduate Engineering Enrollment

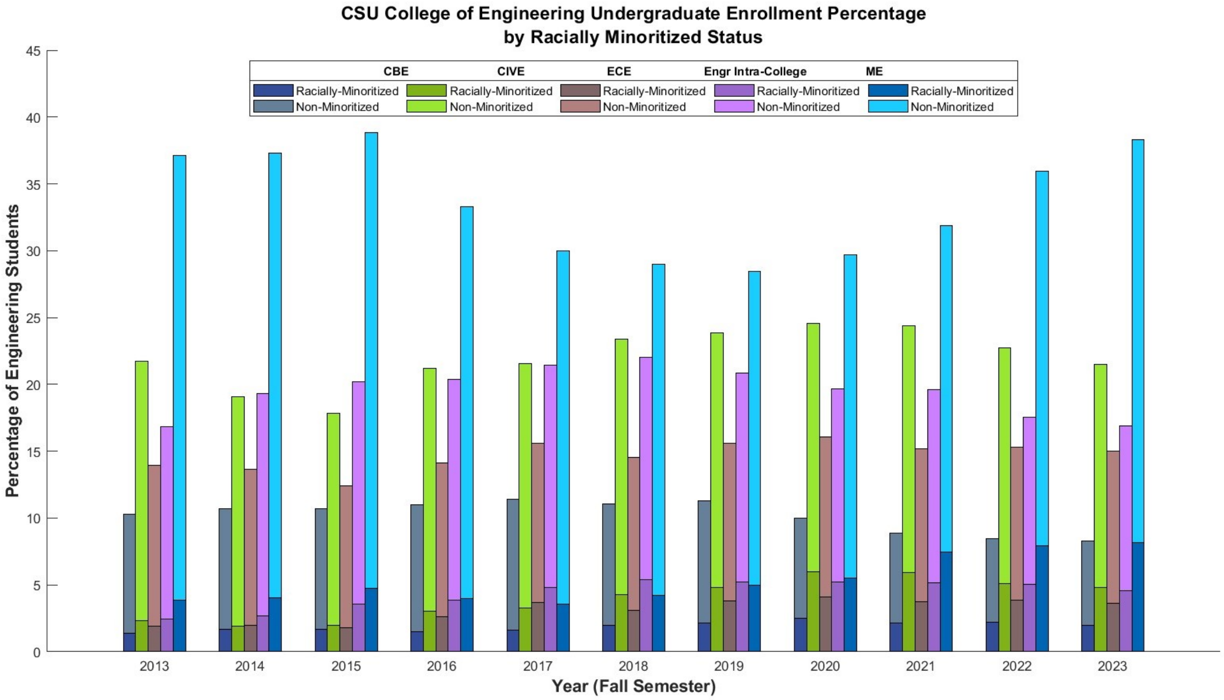


Figure A.1: Percentage of racially-minoritized students in the Walter Scott, Jr. College of Engineering (COE) at CSU from 2013 to 2023, by department. Each bar represents the total number of students in the respective department; the total percentage of students in the department based on the total number of students enrolled in the COE. The percentage for each department is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

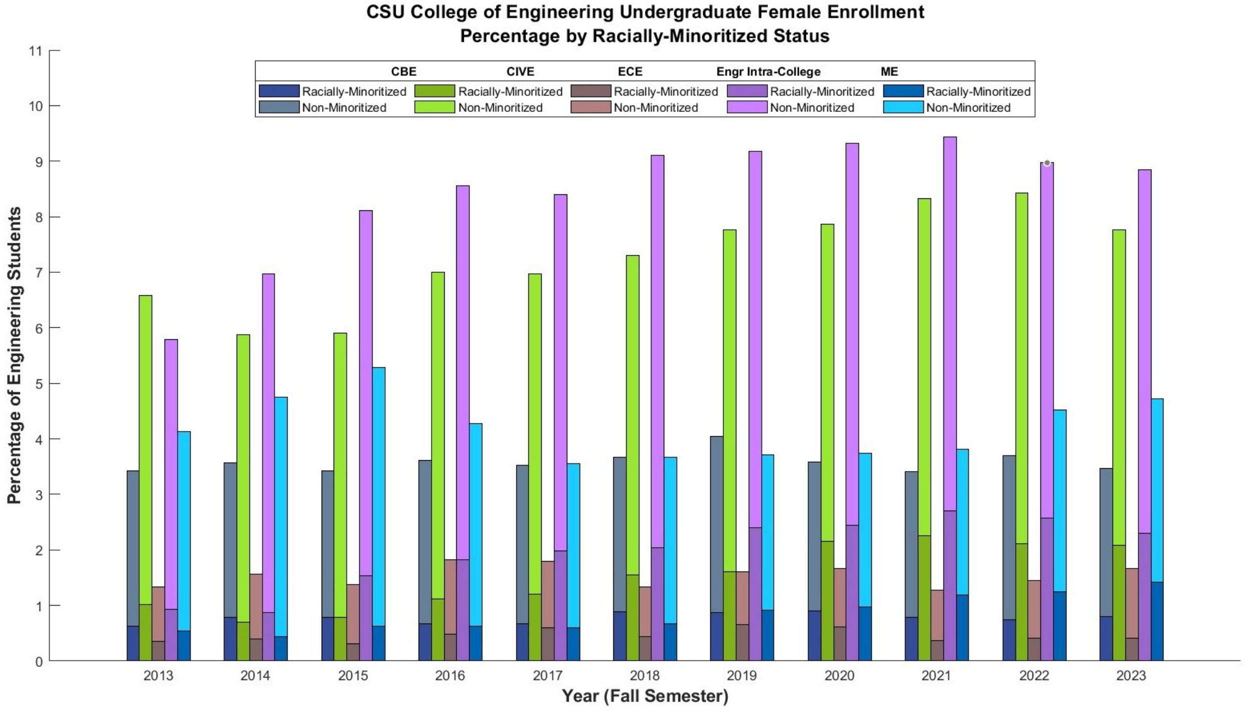


Figure A.2: Percentage of female students in the Walter Scott, Jr. College of Engineering (COE) at CSU from 2013 to 2023, by department and racially-minoritized status. Each bar represents the total number of students in the respective department; the total percentage of students in the department based on the total number of students enrolled in the COE. The percentage for each department is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

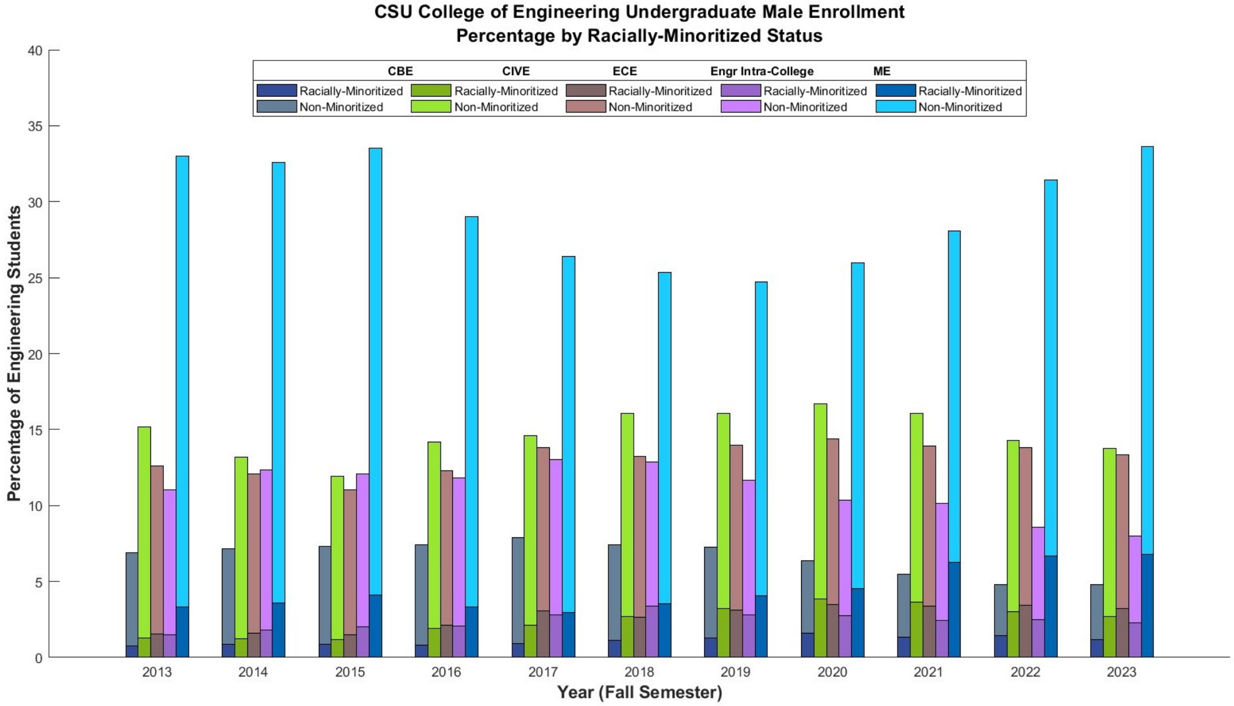


Figure A.3: Percentage of male students in the Walter Scott, Jr. College of Engineering (COE) at CSU from 2013 to 2023, by department and racially-minoritized status. Each bar represents the total number of students in the respective department; the total percentage of students in the department based on the total number of students enrolled in the COE. The percentage for each department is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

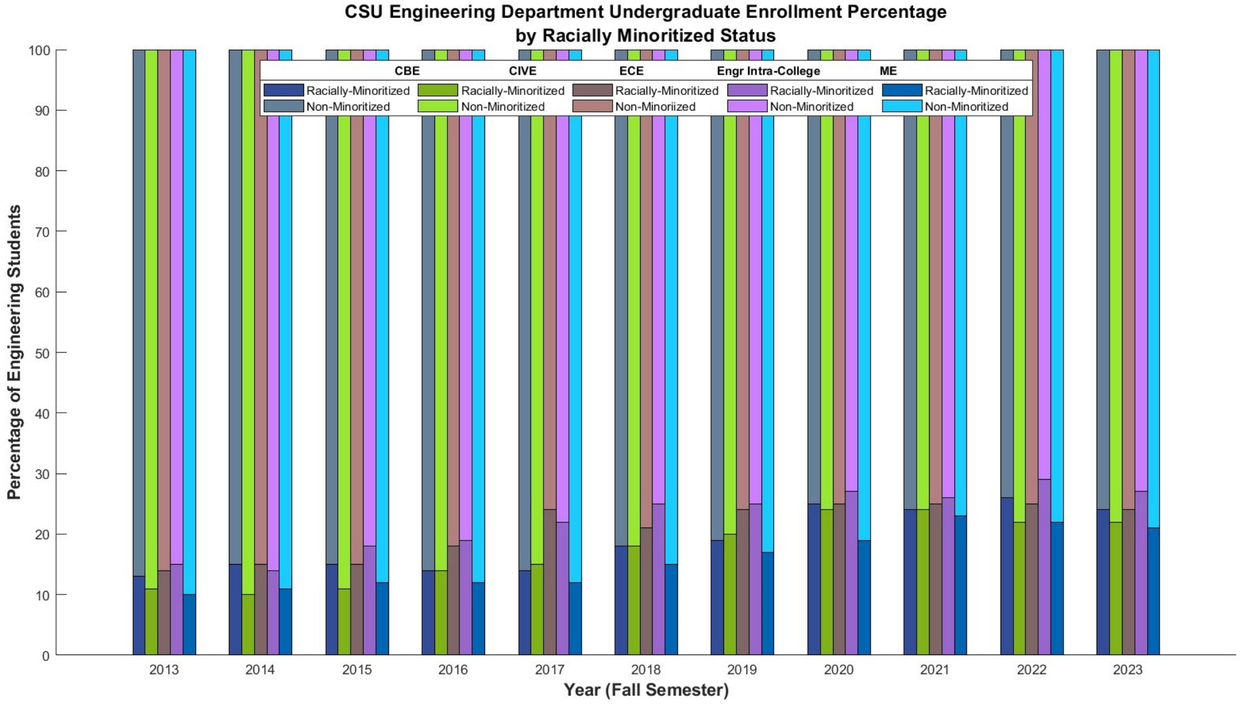


Figure A.4: CSU engineering department undergraduate enrollment percentage from 2013 to 2023, by racially-minoritized status. Each bar represents the percentage based on the total number of students in the respective department. The total (100%) is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

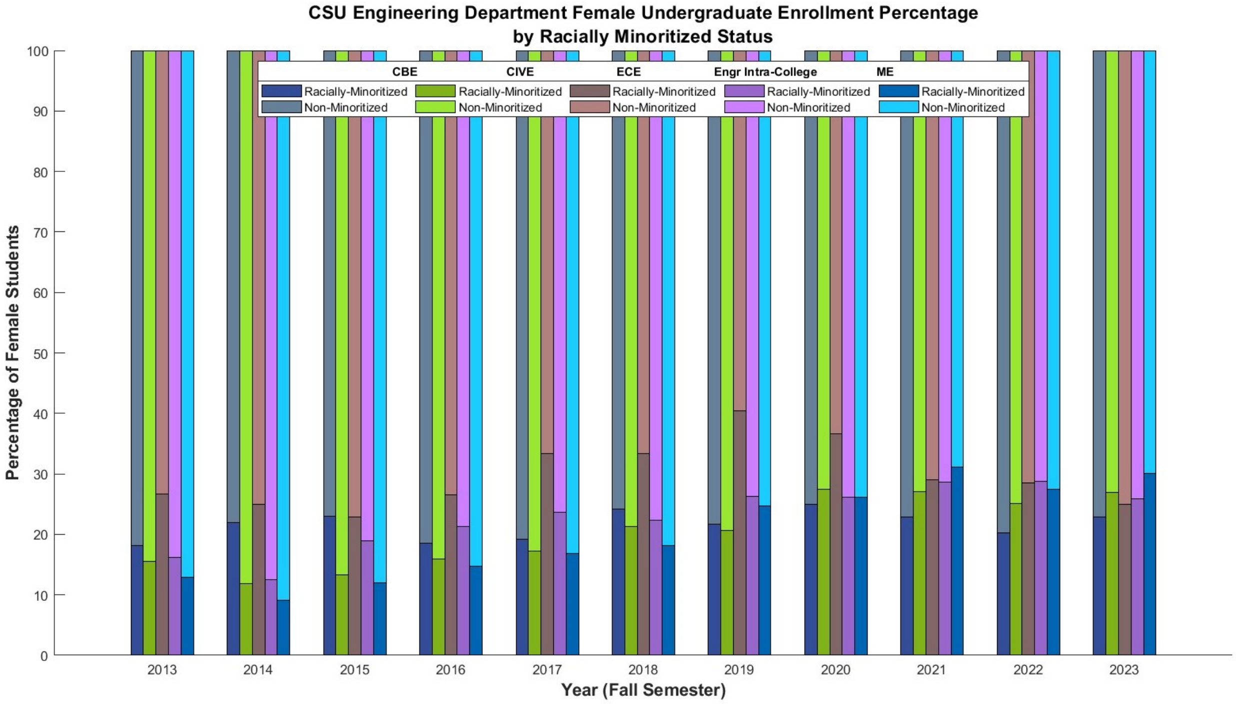


Figure A.5: CSU engineering department female undergraduate enrollment percentage from 2013 to 2023, by racially-minoritized status. Each bar represents the percentage based on the total number of female students in the respective department. The total (100%) is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

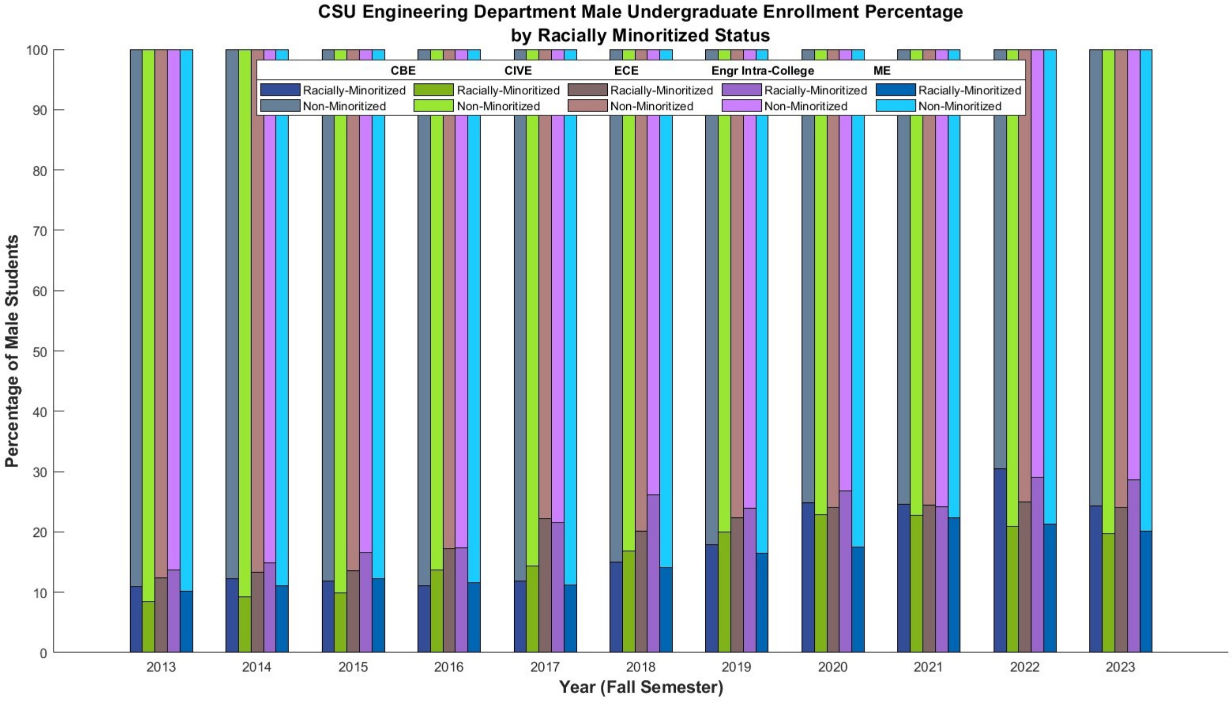


Figure A.6: CSU engineering department male undergraduate enrollment percentage from 2013 to 2023, by racially-minoritized status. Each bar represents the percentage based on the total number of male students in the respective department. The total (100%) is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

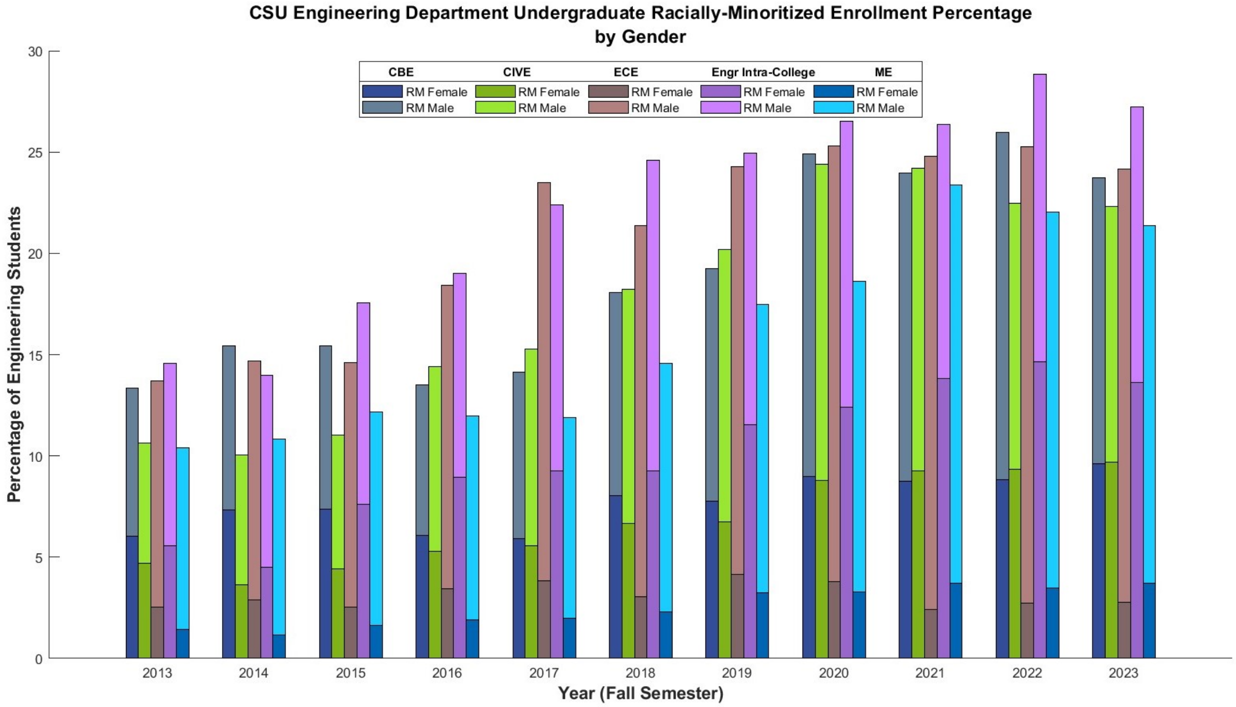


Figure A.7: CSU engineering department racially-minoritized undergraduate enrollment percentage from 2013 to 2023, by gender. Each bar represents the percentage based on the total number of students in the respective department. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

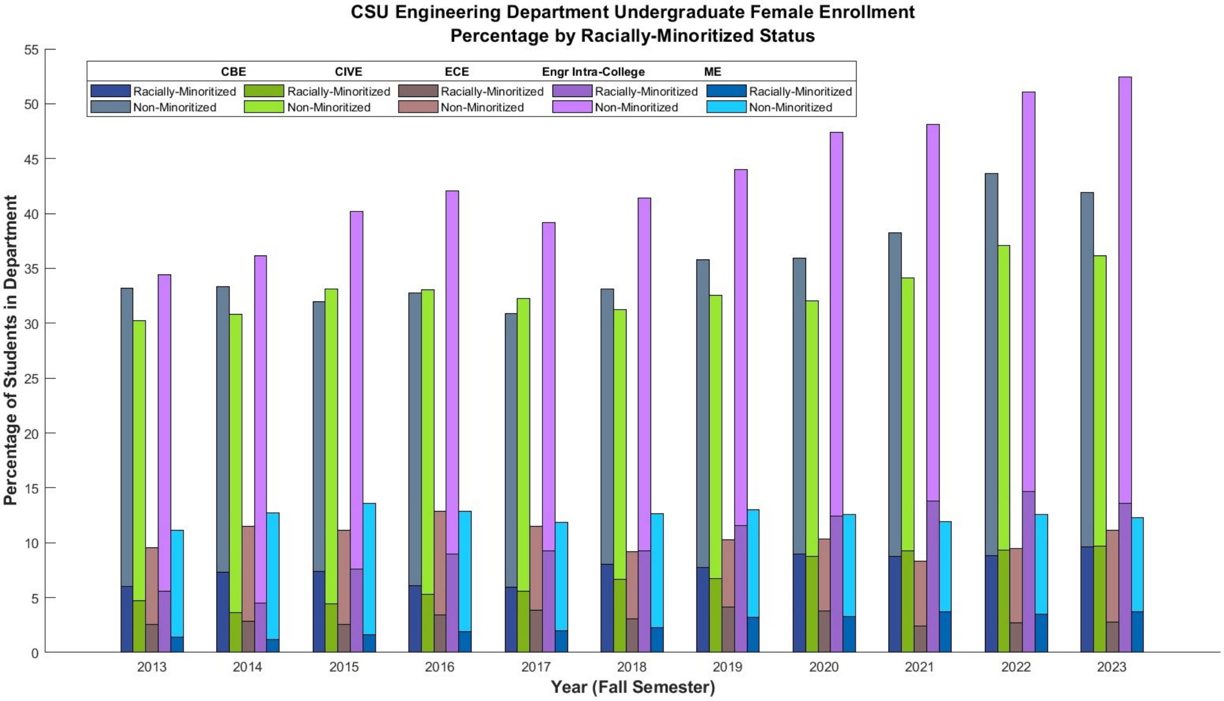


Figure A.8: Percentage of female students in each engineering department at CSU from 2013 to 2023, by racially-minoritized status. Each bar represents the total percentage of students that are female in the respective department; the total percentage of female students in the department is based on the total number of students enrolled in the department. The percentage for each department is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

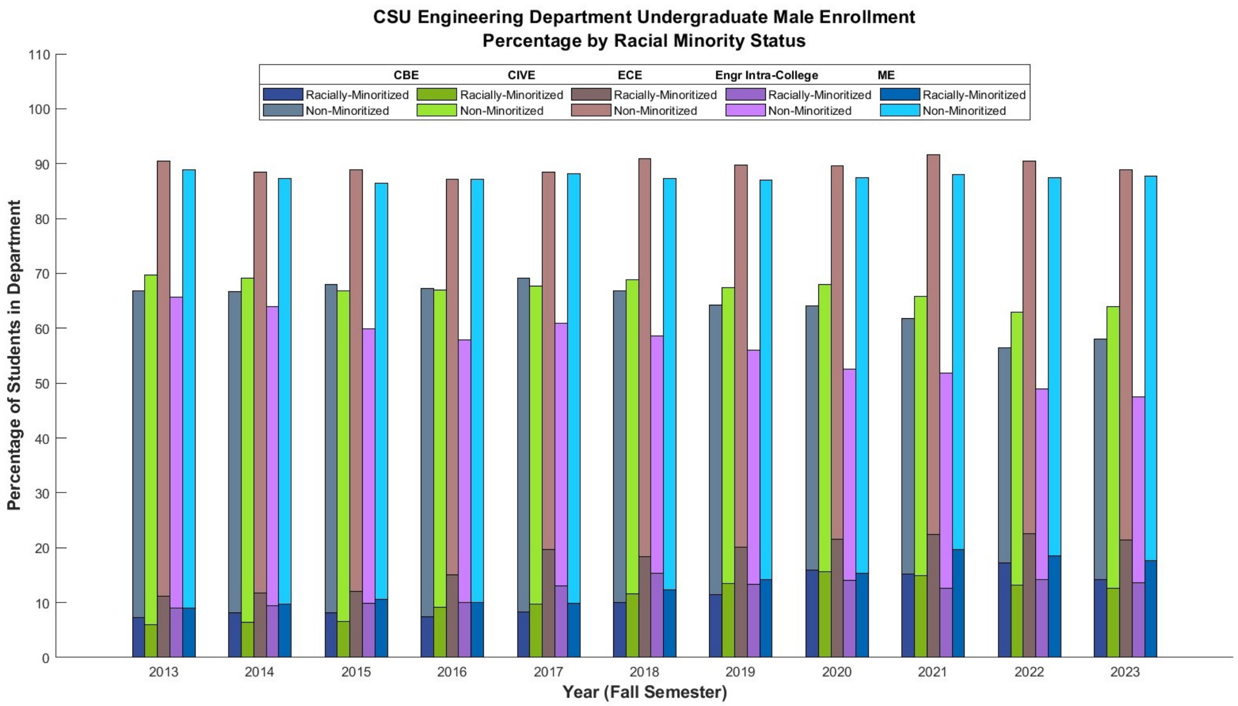
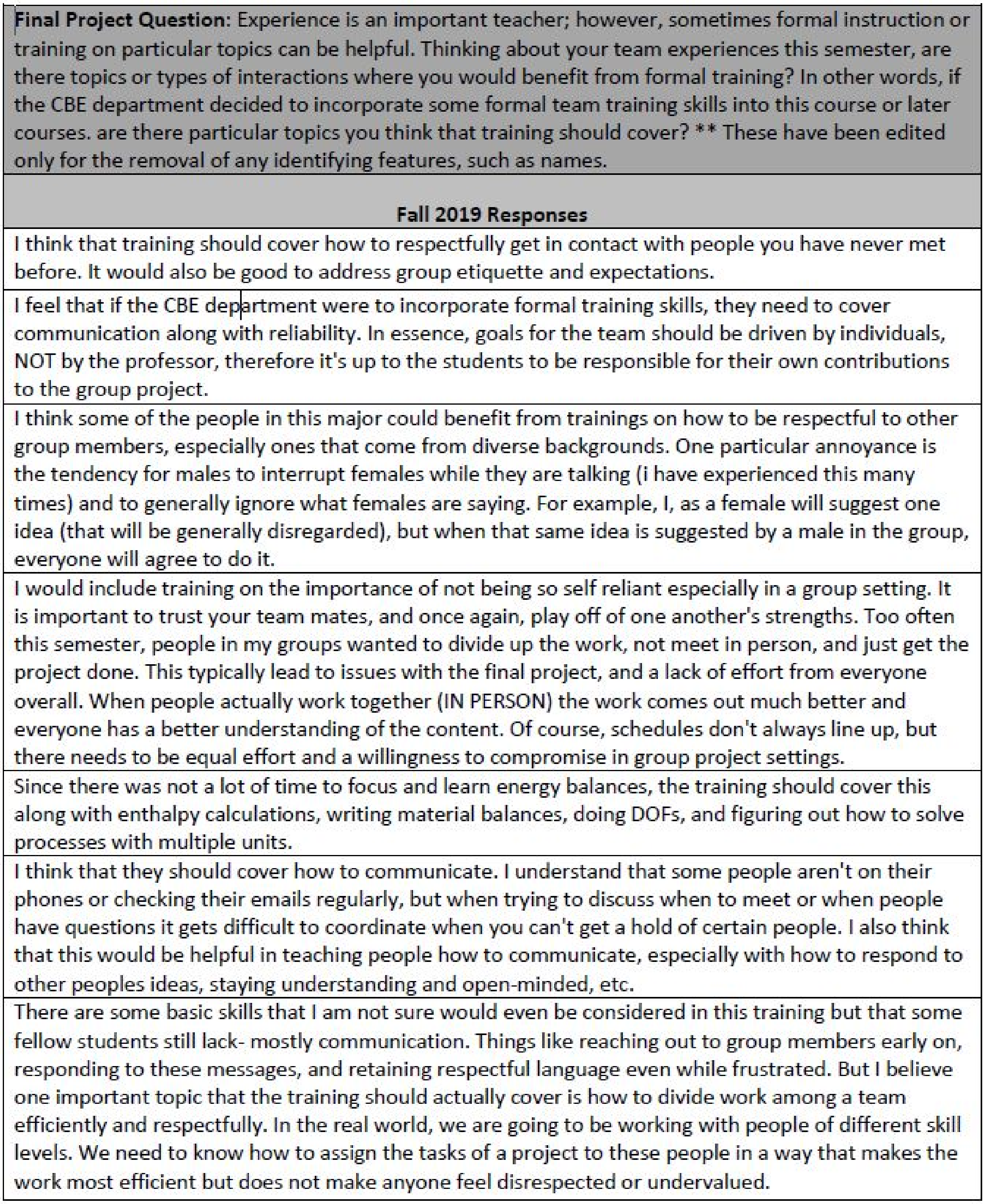


Figure A.9: Percentage of male students in each engineering department at CSU from 2013 to 2023, by racially-minoritized status. Each bar represents the total percentage of students that are male in the respective department; the total percentage of male students in the department is based on the total number of students enrolled in the department. The percentage for each department is divided into the percentage that are racially-minoritized and the percentage that are non-minoritized. In this figure, students in the dual degree program with BME and another engineering degree program are categorized as ENGR Intra-College, along with ENGR Open Option and dual degrees for ENGR Science with Intl Studies, ENGR Physics, Space ENGR, or Teacher Education [196, 198].

Appendix B

CBE 201 Final Reflection Question Responses



Final Project Question: Experience is an important teacher; however, sometimes formal instruction or training on particular topics can be helpful. Thinking about your team experiences this semester, are there topics or types of interactions where you would benefit from formal training? In other words, if the CBE department decided to incorporate some formal team training skills into this course or later courses. are there particular topics you think that training should cover? ** These have been edited only for the removal of any identifying features, such as names.
Fall 2019 Responses
I think that training should cover how to respectfully get in contact with people you have never met before. It would also be good to address group etiquette and expectations.
I feel that if the CBE department were to incorporate formal training skills, they need to cover communication along with reliability. In essence, goals for the team should be driven by individuals, NOT by the professor, therefore it's up to the students to be responsible for their own contributions to the group project.
I think some of the people in this major could benefit from trainings on how to be respectful to other group members, especially ones that come from diverse backgrounds. One particular annoyance is the tendency for males to interrupt females while they are talking (i have experienced this many times) and to generally ignore what females are saying. For example, I, as a female will suggest one idea (that will be generally disregarded), but when that same idea is suggested by a male in the group, everyone will agree to do it.
I would include training on the importance of not being so self reliant especially in a group setting. It is important to trust your team mates, and once again, play off of one another's strengths. Too often this semester, people in my groups wanted to divide up the work, not meet in person, and just get the project done. This typically lead to issues with the final project, and a lack of effort from everyone overall. When people actually work together (IN PERSON) the work comes out much better and everyone has a better understanding of the content. Of course, schedules don't always line up, but there needs to be equal effort and a willingness to compromise in group project settings.
Since there was not a lot of time to focus and learn energy balances, the training should cover this along with enthalpy calculations, writing material balances, doing DOFs, and figuring out how to solve processes with multiple units.
I think that they should cover how to communicate. I understand that some people aren't on their phones or checking their emails regularly, but when trying to discuss when to meet or when people have questions it gets difficult to coordinate when you can't get a hold of certain people. I also think that this would be helpful in teaching people how to communicate, especially with how to respond to other peoples ideas, staying understanding and open-minded, etc.
There are some basic skills that I am not sure would even be considered in this training but that some fellow students still lack- mostly communication. Things like reaching out to group members early on, responding to these messages, and retaining respectful language even while frustrated. But I believe one important topic that the training should actually cover is how to divide work among a team efficiently and respectfully. In the real world, we are going to be working with people of different skill levels. We need to know how to assign the tasks of a project to these people in a way that makes the work most efficient but does not make anyone feel disrespected or undervalued.

If the CBE department were to incorporate some formal team training skills into this course, I think one of the topics that could potentially be covered is how everyone can find different ways to contribute to a project, even if they do not necessarily know the material well.
I do not think that would be helpful, team building practice is not taken very seriously by most STEM majors.
The training should cover how to properly communicate with each other. This could touch on how to communicate with your group mate's ideas in order to properly acknowledge everyone and make everyone's ideals feel validated.
If the CBE department were to incorporate formal team training skills into this course or later course, I think that they should cover the topics of respect and proper reply times/etiquette. One of the main problems that we faced were that people would simply not reply and that was frustrating and sometimes damaging to our ability to complete the project in a timely manner. I think that we should cover the best ways to share coding scripts and how to do that.
I think that the most important team skill that could be taught is personal responsibility, even when in a group. A couple of times over the semester I had group members that weren't nearly as productive (and didn't try to be) as others which didn't positively impact the group.
I think there are topics that can be helpful where small situation with items like m balance elementally and and interaction with it can help underestimating better or Pxy or Txy diagrams where an interaction can help show the point the graphs are trying to show.
I think stressing that engaging teamwork (effort from all tea members) help a project go by faster and more efficiently.
I think that a small seminar based on teamwork could be helpful. In this seminar, students should be taught about organizing meetings, distributing workloads, understanding skillsets, and working as a team. This would likely help people with their groups and group organization.
My first semester I was in Engineering 101 (engineering open option) and they had us do an exercise where we had to establish team rules/goals before beginning a project. This allowed everyone to agree on what to expect from each other.
I think a training on how to actively listen and and ask questions. Often times, people seem to be listening but when it comes time to ask a question it is hard to come up with.
I think that it would be important to be trained on interpersonal skills, implicit bias, leadership, open-mindedness, and construction of objective planning.
I would benefit from formal training on how to interact with anyone outside of my field; I felt that working with engineering students was overall almost entirely positive. I think anyone who was unpleasant to communicate with this semester would never have been hired with that ethic and won't be in the field in the future.
I think that I would suggest the only training necessary would be training groups to evaluate everyone’s strengths and being open to multiple different routes to a solution. This is really important when trying to make the best project possible and finding the best solution.
In particular, I think a course on accurate and complex research would be especially helpful when considering the amount of information we still don't know or when we have to go research a topic to further understand something. I know we already have a coding course but another course where the coding is directly related to research and topics we are studying would be incredibly helpful.
communication and integrity

I think if I was to receive a formal training on a skill from this class I think it would be beneficial to have a training on MATLAB. I understand that we took a course our freshman year that taught us how to use MATLAB, but I found it difficult to recall things that I learned for that program a year ago. I think if a formal training of MATLAB accompanied this class I think it would benefit not only myself, but fellow students as well. As for topics of this training I think it should cover the same topics that CBE 160 covered, but just more briefly and more directed towards what we are doing in this course rather than the beginner skills.
I think scientific communication might be beneficial. It can be difficult to put these challenging concepts into words.
If a training were to exist, it should focus more on how to take and give constructive criticism, so that you're not afraid to add an input when you think someone may be wrong, or take criticism when you are wrong.
I think the training should cover communication skills
I believe that there should be training on group communication. This would better the communication greatly.
I think that the primary formal instruction that would benefit our class, I think that it would be good to cover how to listen and ask the right questions for problems. For example, our instructor could describe a problem, and then we would have to answer series of questions concerning the problem to see how well we actually listened and comprehended the issue. From there, we could be asked to provide questions that, if answered, provided more insight into the problem. This would allow for us to become better listeners and thus better group members.
I think training should cover how to get the most out of working in a team because I think a lot of people have done group projects before but they don't know how to use the team to get the most out of the project.
A topic like leadership training will be useful because I think we always need someone to guide the team to work in the right direction. Even if everyone a team is good, the team still needs someone to connect the work together to make a good report.
Formal team training skills that would be helpful would be how to communicate effectively in a group and job delegation.
I did benefit, although sometimes I felt that I was completely lost and did not do as much as i wanted other times I was the leader in my group and that is why working in groups is so beneficial, because sometimes we might have some difficulties understanding and we can learn from our teammates, or sometimes we really understand the topics and we can explain that to peers that are struggling.
I personally think it's better to learn first hand rather then lean in a class about skills needed for this type of thing.
Team building activities...stuff where we can get to know each other both professionally and informally. We were so busy just trying to get the projects done, didn't get as much of a chance to get to know others to know more about their experiences.
I think that an appropriate topic to teach would be to teach how to divide up work evenly and efficiently. For example how and if work should be done individually or together.
just suggestions in class on each project about how you should break up the project so no single person has to do way more than others
I don't think these skills can be taught. The individual needs to find his own urgency and motivation to complete a task.

In my opinion, teamwork is a topic that can really only be introduced instructionally. Further developing your group work skills requires practice, not instruction. The current incorporation of group projects into the CBE curriculum is exactly what I need to improve as a team member.
no, because I think that particular topics are not that important
That type of training in the classroom should cover communication: teaching people basic how to communicate skills that work in project groups and the real world, time management because for some of my projects that seemed to be an issue for others in my group and I myself sometimes lack there and maybe not always needing to be in charge there were some moments where multiple people in my group wanted to tell others what to do which is okay in some scenarios but its not really conducive to working efficiently.
I wouldn't say it was training due to how they complicated things way to much in the group projects. I would say it was more to weed out people.
Communication skills would probably be the best one to incorporate into teaching. Communication is the most important in any group so everyone is on the same page and no one is lost or lagging behind. Another important subject to cover is how you present yourself since that can either make or break group dynamics at times.
No, these skills are best learned by more interactions with group members. General tips for being a good team member could be helpful for guidance.
I think that 101 should not have a lab. I think the lab for that class is a joke to be quite honest. If 201 had a lab I think it would greatly improve the course work. Honestly I think 101 should be a catalyzed class over summer or like a month consolidated at the beginning in 201. I have a better grade in 201 than 101 because all 101 is is attendance and very simple understandings. Just my opinion. If you want to make an engineering beginning class, make it physics 1 engineer specific. I seriously think it needs to be considered.
We should go over how to allocate work and get things done efficiently so these projects don’t take hours upon hours.
I honestly think strict formal training could be one of the worst things to have and rely upon. I have found that flexibility is the most important part of working in a team, so that would be the only lesson I would want training to impart, however, teaching flexibility isn't exactly the easiest thing to do.
Formal training with regards to rapport-building might be beneficial.
I think that the team trainings should incorporate inclusive diversity due to the fact that engineering is a white male dominated field and people who do not fall under that category should feel accepted and feel as if they van too take on a leadership role in their teams despite being minority.
It might be helpful to take the material and energy balances that we learn in class and apply them into real life situations. I feel like it would be a great way to help figure out why we set the balances up the way we do. Also, it would be good to do that when we learn a big point in class (ex. when pressure/temp/etc. goes up, why does pressure/temp/etc. go down).
Learning to communicate effectively with others is something so important in group projects. I think many people lack this skill and would benefit from training or formal instruction. People should also realize that doing things last minute is ineffective. Time management skills would also be a topic that people may benefit from training or formal instruction.
Maybe teaching some people how to actually put work into a group project?
I think they should cover more on what makes a successful team and how to solve situations that are made difficult amongst classmates.

I feel as if most college students have a general leadership background or experience coming in. Then again, I worked with some group members who definitely did not act as team members. It may be beneficial to train on different types of leadership styles and communication patterns so groups can expand their understanding of each other as they work together. In the case that different group members communicate differently, it may be beneficial for other group members to recognize that they are simply operating differently but still putting in effort.
Yes I think everyone would benefit from a teaching or lesson on communication. I think communication is the most essential part of working with a team. I think if people were trained or informed on the importance of this I believe the team dynamics would be much better.
I think it would be beneficial to have some form of training on how to communicate with a group of people because all of my groups struggled with that this semester. I think the success of all the projects would have been improved if group members simply replied to each other more often, answered questions more clearly, and were able to organize meeting times more easily.
I think there were very few issues in the groups I was in. To be honest, the part of group work I struggle most with is getting the group together. Often teachers leave the students to put themselves into groups by their own choice. I find, for whatever reason that I end up with groups that are less effective than randomly assigned groups. It can be hard to know who would make a good group mate just by sitting near them, so if there was instruction that would help with that, it would be helpful.
No I personally do not think I would benefit from something like that however there are many people who would. I just think it should include tactics on how to communicate and how to deal with issues when they arise.
There should be a smaller class outside of the CBE 201 class where the groups can meet at the same time. Similar to a recitation in chemistry, but with a more practical approach. If there was a class like this, then the students would have a specific time frame where they could work with instructors and their group members. Office hours are similar, but not everyone can make it to those, if there is a class then they can sign up.
The training should cover certain topics such as perhaps the chain of command as some student may have never been in a position of leadership as well as some students maybe have never been a follower. It might be helpful to simulate this and change up who leads in order to train students for real world companies in which you must report to a higher up.
I think that we should do more work overcoming subconscious biases.
Things that I think would be beneficial would be all the characteristics of a bad team/group to act as a foil to understanding how to be a successful group. Also how every single member can be a leader and take initiative.
No I do not believe that is the CBE departments duty unless of course the students took an optional course in which was specifically for teamwork. I'm just here to learn about engineering and science, I have been in the workforce since I was 14 and learned team work throughout all of those experiences. I did not pay to be lectured about team work and the importance of diversity.
I think some training would be good for all students. I think leadership training that covers how to engage all the different types of people to work together would be great. I want to gain the skill and knowledge to be a leader that can bring an entire group together no matter our differences may be. I think the experience of this class helped greatly because it showed me the different types of people that I may encounter. This class also showed me that I am a team player and I want to see my team succeed all together not individual. I think in a later course it would be nice if we had groups where each group member took a turn being a leader trying to engage everyone after being taught some skills in how to do this.

DOF, and Pxy Txy and incorporating all of the related equations for both.
* Group management (Coordination of meetings and whatnot) * How to set up collaborative environments (I set up a google drive folder and shared it with my group each time to keep track of documents) * Teach people to not be afraid to speech up, even if there is a *power* player in the group that wont easily concede the speaking floor. That person's feelings may get hurt, but if they are hogging all the meeting time with their agenda, they deserve to be interrupted.
No the issues in the group project never arose from the team not being able to work together well. The issues that I believe are more important is the minimal amount of formal training that we have received in matlab and the fact that the questions in the group projects were so time consuming to complete.
I honestly think that a training may be a waste of time because know one would take it seriously, i honestly think the best way of teaching us how to work in groups comes from practice, with hard projects which we got in 201.
I don't think a group team work class is needed, it would really do nothing compared to going straight into experience. If there was to be a class, maybe they would try and focus on how to break up assignments for everyone.
I think being trained in a more hands on way with every topic in this course would be extremely valuable. I think a sort of lab would be very beneficial to the student learning experience. I have discussed with lots of my peers and it seems to be the consensus that applying the skills in a real world application or lab format would improve understanding. It could be something as simple as having to design simple systems or reactors ourselves to create a desired output, or essentially just give a physical application to our projects.
I think a general overview of how to work in groups would be good.
Not really, I enjoyed experiencing the work with a team and going to office hours to learn how to do things.
I am not exactly sure what formal training entails and if it would help with learning particular topics. I did think though that this class would benefit from better examples. We seem to have very little knowledge on a topic before having to do it for a group project. If we had a simpler example to work off of it might be easier to figure out what we are supposed to do.
Yes, I think it was very easy to let one person take over the matlab script while you work on other things. Therefore maybe incorporate more matlab in formal training so that we are forced to continue the practice.
If there were to be formal training, the most important thing is communication. Without good communication within the group, even a group with the four smartest people could fall apart. Training should also involve what to do with a team member that hasn't participated yet. While it makes sense that they get a 0 if they did nothing, it does seem better to find ways to incorporate them in a way where they also participate.
I think over the course of a CBE degree, there is plenty of opportunities for a student to practice and refine their group work skills. I believe these are skills that take lots of practice to be good at and I'm not sure a single course would benefit a student so much as developing their way of working in a group.
I think team training skills are very vital skills for this major. Future students can benefit so widely from this if this were to be applied in the CBE department. I believe that learning by doing helps

create so much experience and successfulness for this team training. This allows for more insight into what the future workforce holds by working with more team-based activities, almost most engineering jobs are team-based, which is a very important background to be shown to when studying in the engineering field.
I think the students here are pretty solid on being able to work in a team. I think if any formal training needs to be incorporated, it ought to be in time management skills. Frankly though, the curriculum is probably meant to strain everybody to a certain extent, so the late nights and all that are probably more of a right of passage than something that can be overcome simply by improving on simple time management skills. Also, formal training is also easy to just ignore (for example, if I were pulled into a surprise leadership seminar instead of a lecture about topics relevant to the class, I would probably take out my computer and start working on homework). If it's going to happen anyway though, place an emphasis on open communication. One of the biggest sources of problems I noticed this semester was when messages regarding the group were ignored or forgotten about.
I think training on learning how to communicate your ideas with people that are less familiar with the content that you are trying to express would be helpful. that was a difficult part for me in the group projects. Besides that, I think it would just be important to emphasize organizing times to meet early on after the projects are assigned so that the whole group can actually meet with all of our busy schedules.
I think the department should cover some topics relating to social justice, as it can help build awareness of how to interact with people from different backgrounds and how we can navigate these interactions. It can also build understanding for one another and can help teams build communication skills so we can best work together.
Some of the topics may have required formal training such as how to work together on MatLab code because usually only one person can work on it at a time. Overall, the CBE lab for 101 helped with these teamwork interactions. More examples of how a group should work should be presented initially.
training on how to figure out meeting times and prioritizing meeting times would be helpful
Matlab. There should be a separate class covering the new concepts for matlab and not make us do complex and long matlab on concepts we learnt a week ago and are still absorbing.
I think this formal training should be optional. Many of the people in CBE I have talked to had to do so many team work exercises in high school, in addition to cbe 201, that more training is unnecessary. As stated in the question, experience is an important teacher, and with teamwork, the most important formal instruction is the actual cbe material we would be applying while working in groups. Working in groups cannot be taught. People who are not good at teamwork are often not open to improving and would not take a training like this seriously. I strongly believe the CBE department should focus on improving teaching CBE material instead, especially given that the department has a reputation for poor teaching.
I think that a formal training would be useful to learn about methods for successful communication. That was a major problem in a few of my group projects, so receiving more information on how successful communication in a group can be achieved would be very helpful. Another thing that I think would be good to include in a training are methods of compromise, working with ideas you don't necessarily agree with, and being respectful to every team member's ideas. While these all make sense in a general aspect, sometimes it is difficult to determine how exactly one can implement these things.

Final Project Question: Experience is an important teacher; however, sometimes formal instruction or training on particular topics can be helpful. Thinking about your team experiences this semester, are there topics or types of interactions where you would benefit from formal training? In other words, if the CBE department decided to incorporate some formal team training skills into this course or later courses. are there particular topics you think that training should cover? ** These have been edited only for the removal of any identifying features, such as names.
Fall 2020 Responses
I think if if the CBE department decided to incorporate some formal team training skills into this course or later courses, they should focus on how teams should work together and teach future students to be more accepting to different people and ideas
Training should have a section based solely on responsibility and dependability. Multiple people in multiple groups did not contribute or did not, even when they said they would.
I think that the training should touch more on communication and how to split up the project. Especially being online, it was a lot harder to reach out to people and sometimes group members never even responded or contributed to the final product. The training could also explore different scenarios, such as how to approach a team member who is not doing their work as that is a realistic scenario that students may come across.
I think having a discussion about the bystander effect would be extremely beneficial to a lot of people. I found that a majority of the folks I worked with were definitely plagued by the bystander effect (the belief that you dont need to do a particular thing because someone else will). Each project got increasingly more stressful for me as I realized that many people had this bystander belief and that I was typically the someone else who did whatever did not want to be done. I think a lot of people maybe dont realize this is a thought process they have and bringing awareness would be good.
I would recommend that you would encourage group communication early. And that you would have ideas that would allow for group collaboration and discussions.
I can not think of any, sorry.
I think training should cover communication and how to not come off to other people as "bossy" and a "know it all" because I found that that made me uncomfortable and not want to work on the project with those people. I'm not sure what other skills to train on but I would say that communication is the most important in my opinion.
Maybe how to start working as a team because every time I met with a group for the first time, it was awkward and no one said much.
Yes, I've had plenty of team training through my years of playing competitive sports growing up but that is not the case for everyone in CBE. Maybe some class we have to attend a few times a semester to learn about ways to effectively communicate with each other could be useful.
I would like to dive deeper into more MATLAB problems in applying them to vapor/liquids equilibrium.
Yes, I do. Something that I experienced without almost every group was a sort of unprofessional approach and lack of responsibility. Some would show up to our meetings without any background on the assignment and no pre-work done and therefore a lot of time was spent catching everyone up. So if team training skills could incorporate something about the responsibility of each group member I think that it would really help. Especially for those that show up prepared then end up doing most of the work.

If a formal team training skills course were to be created, I think that a topic that should be taught is encouragement of team members that may not feel comfortable speaking their opinions. I found that within these group projects, there was always at least one or two members of the group that did not speak a word, but just did as they were told. I think that to have a strong group that can collaborate effectively, such people must gain the confidence to actively participate.
Yes, the department should incorporate brainstorming strategies and a methodology to solve problems as a group and not only by ourselves
I dont think forcing engineers to talk to eachother or build team mentality face to face this early is a good thing. It would add stress to those who are antisocial to begin with. Maybe in the next CBE course that would be a good idea.
I think this experience is sufficient, but if there was training it should include anger management, problem resolution, time management, the physiology behind seeking help when necessary, and good listening skills. I also think training in leadership would be beneficial to work better in a team.
Perhaps general insights into how we're going to be applying these CBE skills we've been learning to the real field would be helpful to better prepare graduates to hit the ground running when they both search for and land their future careers. This isn't too dissimilar to how it would be helpful earlier in high school to incorporate practical economics (bills, taxes, loans, payments and the works) into the curriculum, as knowledge on these subjects would be useful to the graduating seniors. All in all, practicality is king in my opinion, and if we can get guidance or even training on precisely how we could be applying what we've learned to our future jobs, that would be best.
I would love to have more MATLAB training with stuff that was not covered in CBE160.
If a formal team training was incorporated into the course, the CBE department might benefit from lessons in how to approach difficult problems as a group (helpful strategies for working together) or by clearly explaining what is expected by the instructor for these projects.
The training should definitely include how to best contact team members, especially those that are unresponsive. It's hard for me to ask others to do a part of the project if they don't volunteer because I feel like I am asking for too much. If we had more training on how to better approach this and resolve conflict and uneven effort within groups I think my groups would have functioned much better.
Not that I can think of.
I can't think of any particular topics that would really benefit from formal team training.
Formal training regarding collaboration techniques and how to deal with people would be helpful.
I feel we could potentially benefit from a lab in this course if it correlated with the processes which we have to understand and analyze. For instance, being able to see a separator in action and visually see how the different species and streams are separator would definitely be helpful to some people. Being able to operate or see how a reactor works would also be a good topic to incorporate in formal training.
I would want some team building activities that didn't need to be a group project instead going out and playing games. That way it is less stressful.
I think some main topics to cover would be organizational skills, and social anxiety in a group setting, and how it may affect people in your group.
I think the training should cover sexism, racism, and all other discriminatory things that happen in the workplace. While this year, I have not seen any of these things happen, I know that they do and I think everyone could benefit on ways to combat this and expand their knowledge on how to be inclusive.

I feel team skills are developed naturally and the more group projects we work on will expand our team skill the fastest
Project planning. The process of dividing up work and setting deadlines is absolutely essential to any project working smoothly. In addition, setting up avenues of contact is necessary, because the number of times that someone missed a message or didn't get a memo was high, so making sure that's clarified and formally taught seems useful.
I think the most important aspect of a group project is communication.
None, I think the group projects are enough training. I have experience working in several fields before and those kind of formal teamwork training always created animosity and never had a noticeable effect. Teamwork is something to learn naturally through practice.
- Accepting mistakes - How to disagree with explanations, really focus on not coming off as condescending - Communication, communication, communication Actually, just think of the stereotypical training and that's where I was going. The only issue with that is that it generally makes no difference in how people feel and act as it is really, really hard to pull off well. Anytime in my life, albeit coming from high school and prior, there has been an attempt to formally teach team skills it has changed nothing except be mocked relentlessly. I know that this was less of the question that was asked but it would be my main concern with implementing formal training for team skills.
I'm not really sure which topics I think the training should cover. I would like more team training with MATLAB because some people are better at coding than others. I personally am not the best at coding. I think it would allow people to gain better team skills.
I think it'd would be beneficial to have designated class time to work on them, when an instructor is present, to help with some of the more confusing areas.
I think this training should be over any learning topic in the class that has a very applicable principle to most things we will see in our future careers.
I think the CBE department is training us in a well-rounded way. The courses are building off of each other and they are creating a solid foundation for later courses. I think that more MatLab training for this CBE 201 incorporated into the lectures would be beneficial or having an advanced MatLab course to come after CBE 160 would be very helpful. Overall, the courses seem like they are going to prepare me well for the work field.
I think organization within a group would be a big topic needed to be covered. I think that was a big issue throughout, was just being able to organize all the information and organize to give us the ability to effectively meet.
It would be helpful to have training on how to handle group members that do not contribute to projects. It would also be helpful to have training on how to assign tasks equally and how to contribute when everyone has different strengths and weaknesses. The best way to learn is from experience; although this class has definitely given us experience in teamwork, some people were still not contributing by the end of the semester. Therefore, it may be helpful to have training in teamwork.
There most definitely needs to be formal instruction with applying the CBE knowledge that we learn to matlab. I feel like several times this semester I understood the content but had no idea how to

apply it to matlab. For loops and fsolve functions should be reviewed because I think a lot of engineering concepts require that knowledge to apply them.
If this were to happen I think one topic or type of interaction that should be covered is time management or something along those lines. Some of my groups were totally fine and got together great and for others, it was more difficult to get someone to show up or a time that didn't work for everyone. That made the projects to be really difficult and I think for a lot of people it wasn't only showing up to group discussions but also doing their part. Lots of people, in the beginning, did their part of the project very last minute and I think it was because they didn't know how to time manage and I do think it's a skill people should know.
I think that would be a really great idea! I think that they should include lots of information about how to communicate within a diverse group of people and how they may think differently from you, and teach us about how to see from different perspectives.
If there was any topic of training, perhaps it might be good to incorporate some team building activities. These activities may increase communication, and make it easier for others to work together.
I think formal skill training could be useful for some groups. With this semester, whenever a group member would not show up or do the project, it was hard to contact them because of the inability to meet in person at class. I think most of the issues in my team experiences stemmed from something like that as opposed to formal training. If there were a formal team training skills unit in a course down the line, I think that the most useful tool to me would be learning how to correctly organize a group and divide up work in order to work most effectively.
All Matlab code I think I would need formal training on as a lot of my peers had very little clue on how to apply it to the projects and a lot of guesswork was used. Also being able to work on a code with multiple different people needs formal training as it is hard to be able to understand what everyone is doing and why especially when working on harder and longer projects.
I think training should cover how to properly approach a task when working with a larger group, especially when a project head is not established. The content was covered properly, so I think training would be most effective if it covered the dynamics of teamwork and how to accomplish a goal.
I think it should cover how to engage people who don't want to help and how to handle situations like that because there is nothing more frustrating that when people don't help and it causes people to be less open which makes the issue worse. It would be good to learn strategies on how to handle that.
I don’t think there are any topics that need formal training. If a topic had to be covered I would suggest how to be respectful when someone disagrees with you or two members of the group are disagreeing.
I think overall some form of formal training could be nice, not for a specific topic but so that people learn to work in groups. Over the course of the year there were so many people I was supposed to work with that never contributed, maybe this is because of laziness but I was thinking that maybe they just don't know how to communicate with groups.
I think training in doing in your own but matches the description that the group agreed on.
The more formal training the better, the more hands-on we get with material and topics relating to CBE the better since it will help in developing my experience.
I would like to be more confident when it comes to team training. So every activity that involves getting the students out of their comfort zone will benefit myslef.

I honestly think incorporating "formal team training" might be counterproductive in the long run. I know not everyone is going to be inherently good at working in a team, but just giving students enough opportunities to work in a group is the best training you could give them. I don't know if there would be enough valuable information about working in teams that you could give that students can't learn just from experiencing a group collaboration to warrant adding a whole class for it.
In my opinion, formal team training would be extremely beneficial. Although I have seen a lot of great collaboration and teamwork in this class, the skills that my peers, and myself, have regarding teamwork could be refined with a training session.
There aren't any particular training skills I can think of off the top of my head, though it would help to be taught how to communicate with a member you may think is lost.
I believe formatting for formal presentations as well as presenting more projects will help boost the greater experience for presenting in the workforce.
I'm not sure if team training skills can be taught except by actually being part of a team. Unless the training course told individuals where their faults are, I don't see it helping much. If the department decided to incorporate some team training skills into the course(s), then I think being an active member is a big topic. It doesn't mean they have to talk, but to participate and give their voice. Most of the groups I worked with had one or two people who just waited until a task was given, completed the task, and then sat on the sidelines until the project was submitted.
I personally believe that online differs greatly from in person group projects. However, I don't think any one thing should be formalized in a group project. All types of people work differently with each other, with ranging skill sets, and peoples desire to complete the project. As time goes on, students will have to learn that having on or two members not present can drastically effect success.
The training should cover how to approach a problem when given a big task. There seemed to be a lot of confusion of how to start the projects, so having someone with experience discuss their approach to a difficult task could be helpful. There was a lot of information and we did not always know what to do with that. I think exercises where we are given a topic, have to determine what we need to do and how to execute that plan would be helpful. There just seemed to be a lot of uncertainty in the approach when we have not seen many real world problems in engineering before.
I think it would be beneficial to incorporate communication training. I think it would help teams communicate better and encourage people to share ideas within the group.
For training, I think it is very important to learn communication and how to speak opinions when in a group setting, as well as equality being a major part of engineering right now,
I think the training should cover respect of others in a group project environment, as well as even splitting up of work and the importance of doing your part. This is all basic stuff, but so important.
I have no idea, really. I think people, at this point in their academic careers, have done an immense amount of group work, so should have pretty good idea of how to work with other people.
I think accountability should be covered as there were way to many people who very much had the attitude of someone else will do it. I also I think this training should cover the importance of coming prepared and being ready to participate.
yes, I think training students to interact with a diverse group and teaching them to accept everyone for who they are, giving such type of training would change a lot of people life. this is because for our community to be strong we need a diverse group, as an engineer to solve a world problem we need a heterogeneous Engineer not Homogeneous, so we need to teach everyone to accept each for whom we are.