Genetically-Programmed Artificial Cells and Multi-Cellular Machines

Richard M. Murray
California Institute of Technology

Submitted, Vannevar Bush Faculty Fellows Program
9 January 2017

Abstract

The long term goal of this research is to create genetically-programmed artificial cells and multi-cellular machines that can carry out useful engineering operations. Unlike more traditional synthetic biology approaches, artificial cells are non-living: they make use of genetic elements provided by biology, but they do not replicate, mutate, or evolve. Applications range from synthesis of bio-compatible materials, to environmental monitoring and remediation, to self-assembly of complex multi-cellular machines. Pursuing this goal requires fundamental research in biological engineering, aimed at moving from creation of clever biomolecular devices to systematic specification, design, integration, and testing of circuits, subsystems, cells, and multi-component systems.

The specific emphasis in this proposal is on establishing the design methodology, subsystem technologies, and system architectures that can enable the development of genetically-programmed, artificial cells. The proposed effort emphasizes demonstration of the fundamental principles that can enable construction of complex, genetically-programmed systems, leveraging work by other groups on various components and subsystems. DoD applications include production of specialty materials, field-based conversion of biomaterials, detection of molecular signatures and toxins, and environmental monitoring and remediation. A major element of the proposed effort is dissemination of results to DoD and other potential users.

Our technical approach will include a combination of theory, computation, and experiments, aimed at developing a scalable and modular framework for creation of biomolecular circuits and systems that implement complex and robust behaviors. Central elements of the approach will be the use of control theory as a unifying mathematical basis, the use of feedback as a core mechanism for managing uncertainty and design of dynamics, and the use of cell-free methods for prototyping and implementation. System-level demonstrations will be used to demonstrate the feasibility of artificial cells and multi-cellular machines, validate the proposed design framework, subsystem implementations, and assembly techniques, and identify areas of future research.

Richard M. Murray, California Institute of Technology

Statement of Objectives

The long-term goal of this research is to develop the methodology, components, subsystems, and system-level architecture required to design and deploy artificial cells and multi-cellular machines. DoD applications include production of specialty materials, field-based conversion of biomaterials, detection of molecular signatures and toxins, and environmental monitoring and remediation. Cell-free methods offer many advantages to the DoD, including portability, orthogonality, safety and stability.

We do not anticipate that a single project can complete such an endeavor in 5 years time, so the focus of this project is on the fundamental research required to demonstrate its feasibility. The specific objectives for this project are to:

1. Develop a mathematical framework that allows modeling, analysis, and design of complex biomolecular feedback systems, ranging from core biomolecular processes, to component-level dynamics, to subsystem behaviors, to multi-subsystem interactions.

2. Implement and characterize a collection of at least three biomolecular subsystems capable of providing robust and modular operations in spatially-isolated, cell-free environments.

3. Create and optimize a method of assembling multiple biomolecular subsystems into a multi-element ensemble with structured interactions between the individual subsystems.

4. Demonstrate feasibility of artificial cells and multi-cellular machines by implementing at least two experimental demonstrations consisting of interacting, multi-subsystem behaviors that validate the proposed design and implementation frameworks and help identify future of areas research.

5. Disseminate the results of the project via community-driven, open-source software and wetware repositories, as well as collaboration with DoD researchers in synthetic biology.

Our technical approach will include a combination of theory, computation, and experiments, with the core fundamental research aimed at developing a rigorous, systematic approach to the design and implementation of robust, modular biomolecular circuits and subsystems that can be interconnected to create a variety of complex behaviors. Central elements of the approach will be the use of control theory as a unifying mathematical basis, and the use of feedback as a core mechanism for managing uncertainty, implementing desired dynamics, and providing modularity that enables successful design and demonstration of multi-subsystem behaviors. Experimental work will focus on the use of cell-free methods for prototyping and implementation, building on current and prior work at Caltech as well as that of collaborators and the broader synthetic biology community.

The proposed project will be carried out by a combination of 4–6 graduate students (with 3 supported by the grant, the others on external fellowships), a full-time technician, and undergraduate researchers (up to 6 each summer, plus academic year projects). We anticipate active collaborations with other research groups—at Caltech, at other universities, and across the DoD—as well as particiaption by visiting researchers to help carry out the broader vision.

Research Effort

1 Motivation and Vision

Synthetic biology has made significant strides over the past 10–15 years in demonstrating the ability to engineer biological systems by "programming" DNA to carry out specific operations both in cells [31, 39] and in cell-free systems [26, 60]. The currently demonstrated systems have been of modest complexity (typically less than a dozen programmed elements) and have focused on relatively simple operations (oscillators, logic operations, metabolic pathways). A major challenge in the field is learning how to systematically design and implement biomolecular circuits of much higher complexity (hundreds to thousands of programmed elements) that can carry out more complex operations, spread across multiple functional units (*a la* multi-cellular organisms).

One approach to moving the field forward is to shift the focus from engineering of living organisms to the creation of genetically-programmed *artificial* cells and multi-cellular machines. Unlike more traditional synthetic biology approaches, artificial cells are non-living: they make use of genetic elements provided by biology, but they do not replicate, mutate, or evolve. Cell-free systems offer many intrinsic advantages, especially for the DoD, including portability (e.g., paper-based cell-free circuits [47]), safety (via true orthogonality and lack of self replication), and stability (due to lack of mutation and evolution). In this proposal we define a set of fundamental research challenges related to the design of artificial cells and multi-cellular machines that present a high-risk, high-reward approach to synthetic biology and its applications.

A schematic diagram of the type of system that we envision is shown in Figure 1. An artificial cell consists of a number of subsystems, described in more detail below, that provide the core functions required for operation in both single cell (left) and multi-cellular (right) environments. We propose to carry out the fundamental research required to demonstrate that systems of this complexity can be designed and implemented.

Project goals and objectives We propose to develop and demonstrate the key design tools, molecular components, and system-level architecture for artificial cells and multi-cellular machines. While we do not anticipate that a single project will lead to a functioning artificial cell or multi-cellular machine in in 5 years time, we believe that it will be possible implement and integrate a variety of subsystems that establish the feasibility of artificial cells and serve as a starting point for a larger community effort.

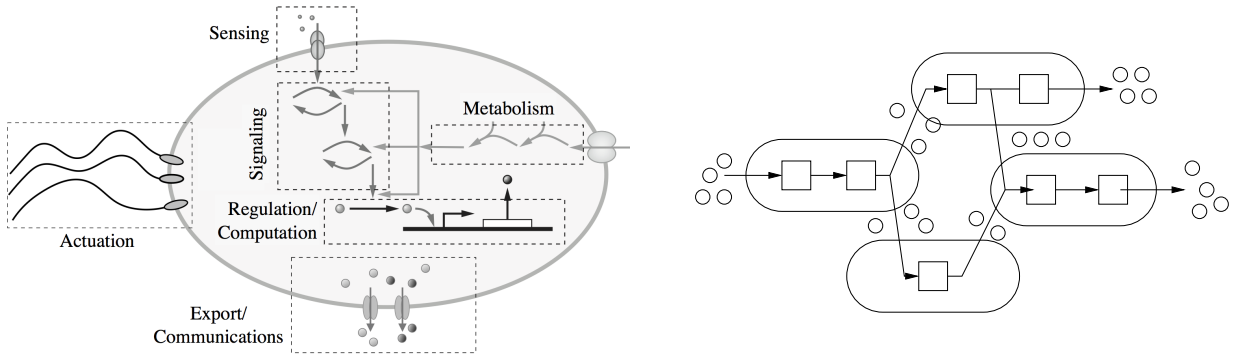


Figure 1: Conceptual diagrams of an artificial cell (left, adapted from Del Vecchio and Murray [12, Fig. 1.5]) and an artificial multi-cellular machine (right).

Toward that end, the specific objectives for this five year project are to:

1. Develop a mathematical framework that allows modeling, analysis, and design of complex biomolecular feedback systems, ranging from core biomolecular processes, to component-level dynamics, to subsystem behaviors, to multi-subsystem interactions.

2. Implement and characterize a collection of at least three biomolecular subsystems capable of providing robust and modular operations in spatially-isolated, cell-free environments.

3. Create and optimize a method of assembling multiple biomolecular subsystems into a multi-element ensemble with structured interactions between the individual subsystems.

4. Demonstrate feasibility of artificial cells and multi-cellular machines by implementing at least two experimental demonstrations consisting of interacting, multi-subsystem behaviors that validate the proposed design and implementation frameworks and help identify future of areas research.

5. Disseminate the results of the project via community-driven, open-source software and wetware repositories, as well as collaboration with DoD researchers in synthetic biology.

Contributions to fundamental research This project involves a combination of research activities that will combine existing capabilities into system-level functions, leverage the work of others (including collaborations), and develop new theory and methodologies motivated by integrative technology demonstrators. We anticipate fundamental research contributions on three fronts:

- New systems-level frameworks for synthetic biology research: the use of mathematical modeling, analysis, and design tools as an integral part of systems-level design will provide new theory and algorithms for biomolecular system design, applicable to both cell-free and cell-based systems.

- Development of novel biological circuits and subsystems: the individual circuits and subsystems that will be designed and implemented will provide new examples of engineered biological systems, of higher complexity than current circuits, with applications relevant to national security.

- New insights into multi-component synthetic biological systems: the integration of multiple subsystems in support of artificial cells and individual cells as part of a multi-cellular machine will enhance the ability to create more complex systems from biological components.

2 Proposed Approach

By leveraging work in the synthetic biology and molecular programming communities over the past decade, we are plausibly within 10–15 years of being able to produce genetically-programmed artificial cells and multi-cellular machines that can carry out useful engineering operations. Pursuing this vision will require new approaches to biomolecular systems engineering, focused on moving from creation and characterization of devices and simple circuits to systematic specification, design, and integration of circuits, subsystems, cells, and multi-component systems.

A high-level diagram of the type of system we have in mind for a single cell is shown in Figure 1. The architecture consists of a collection of subsystems, most of which we believe can be implemented using available technology. At this stage, it is not clear whether the artificial cell should be "booted up" using cell lysate with genetically-programmed DNA sequences or whether subsystems should be created separately and integrated in fully expressed form. In either case, we imagine that the artificial cells would not replicate themselves per se, though some self-assembly would be required. They would only function as long as an external (probably chemical) energy source is present.

Multi-cellular systems would consists of a collection of individual cells that interact with each other through physical and/or chemical interactions, as illustrated in Figure 1 (right). Each cell would consist of one or more subsystems that carries out part of the work of the overall system, with the ability to import and export biomolecules that can serve as either substrates for processing or communication channels. The segregation of individual functions into separate cells will enable more modularity and higher complexity behaviors.

2.1 Cell-Free Synthetic Biology

The work that we propose builds on a set of tools for cell-free synthetic biology that my group and others have developed over the last 5–10 years. These efforts involve the development of systematic frameworks for implementing circuits and pathways [17, 21, 23, 25, 40, 42, 57, 70, 73], methods for interconnecting components and isolating unwanted interactions [16, 23], methods for compartmentalizing circuit operations [1, 8, 46], and methods for spatially localizing molecules using programmable scaffolds [28, 68]. This section provides a brief overview of these techniques.

Genelet circuits "Genelet" circuits are DNA- and RNA-based systems that rely only on transcription and binding of complementary sequences of DNA and RNA to create genetically-programmed functions [33, 35, 36]. The primary mechanism of action in a genelet circuit is the use of a partially double-stranded sequence of DNA with an incomplete promoter region that can be completed by using a single-stranded DNA activator. Upon activation, RNA polymerase transcribes downstream region into RNA. Activators can be displaced from the template if they present an overhang, or toehold: this exposed area allows an inhibitor strand (RNA or DNA) to initiate binding and eventually strip off the activator from the template to reach a more favorable thermodynamic state. In addition to this basic mechanism, sequestration and degradation reactions can also be utilized.

Figure 2 shows an example of a oscillator designed using genelets [16, 36]. Other circuits that have been constructed include a "rate regulator" circuit in which the production two RNA species are modulated to maintain a constant ratio of production rates [17] and an "insulator" that is used to minimize the coupling between an RNA-based oscillator and a set of DNA-based "tweezers" that opens and closes based on the presence of a complementary RNA-strand [16]. Genelets provide a highly programmable approach to implementing transcription, repression, sequestration, and degradation reactions. They can operate in some transcription/translation systems (e.g., the NEB PURExpress system) and we are exploring their use in cell-free extracts.

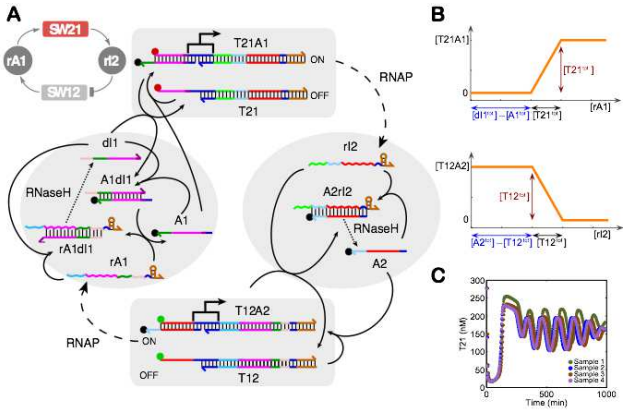


Figure 2: RNA-based "genelet" oscillator [16, 36].

Cell-free prototyping Over the past five years, we have helped improved on a transcription-translation (TX-TL) platform—originally developed by Vincent Noireaux [56, 62]—to construct and characterize synthetic gene circuits in a cell-free environment. The TX-TL system uses *E. coli* extract containing the cell's protein synthesis machinery and can express genetically-encoded circuits by simply adding DNA encoding the desired circuits into a test tube. Importantly, TX-TL can use linear DNA from a PCR machine, enabling rapid and inexpensive prototyping.

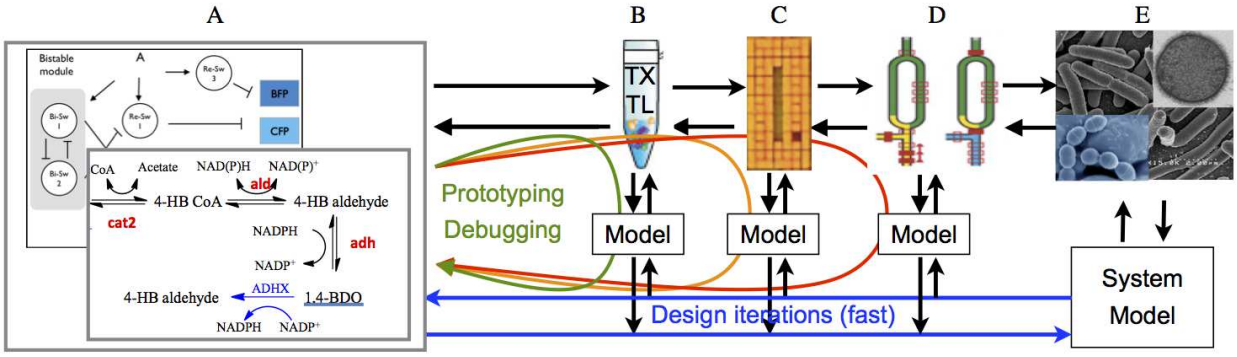


Figure 3: Biomolecular breadboards framework for rapid prototyping of biological circuits.

An overview of the current biomolecular breadboard technology is shown in Figure 3. Column A represents examples of circuits that we have designed and implemented, including a biomolecular "event detector" [28] and a pathway for producing 1,4 butanediol [70]. The main component of the breadboard is the TX-TL extract system, represented in Column B. We typically run 10 μl reactions and are able to obtain 6–10 hours of gene expression and circuit/pathway operation. A detailed modeling toolbox is also available for simulation and analysis of circuits and pathways [65]. More sophisticated test environments are shown in columns C and D, which illustrate the use of a commercially-available droplet-based microfluidic system and a continuous flow reactor (development by EPFL [41, 42]). We are able to transform circuits into *E. coli* (column E) and are currently developing extracts for other micro-organisms.

The cell-free toolbox is a significant advance over commercially available systems, both because of cost (10-100X cheaper) and because if provides a more realistic prototyping environment. We are able to prototype circuits with a cycle time of less than 8 hours from design to data [42, 63]. We have successfully used TX-TL prototyping to implement regulatory circuits [23], decision-making logic [21, 25], novel genetic oscillators [42], and metabolic pathways [70, 40, 73]. We have also developed and published improved methods for using TX-TL as a prototyping tool [62, 64] and as an educational resource [24, 25]. Using a recently acquired, DURIP-supported, acoustic liquid handling system (Labcyte Echo), we are able to test hundreds of circuit and pathway variants at a time, enabling high throughput data collection and design space exploration.

An example of the design workflow for TX-TL-based circuit design is shown in Figure 4. Starting at the left and proceeding clockwise, the workflow begins with a design concept and mathematical model of the desired circuit. Individual components are characterized and tested, then combined for circuit-level testing and characterization. Finally, the entire circuit is assembled onto a single plasmid for final validation. This design-build-test cycle can be iterated multiple times, if needed. In this project we will adapt this workflow to focus on artificial cells and multicellular machines, with the *in vivo* implementation replaced with an artificial cell implementation.

Characterization of biological circuits Cell-free systems also play an important role in our ability to develop high fidelity models and characterize individual biological components, whether for future cell-free usage or for cell-based circuits. Examples in my group include characterization of genetic context effects due to supercoiling [71], occupancy effects in integrase circuits [3] and effects of resource limits on circuit performance [57]. These results all rely on a combination of modeling, analysis, and (cell-free) experiments. In addition, we have developed open source modeling toolboxes that capture many of the important details of cell-free systems [65].

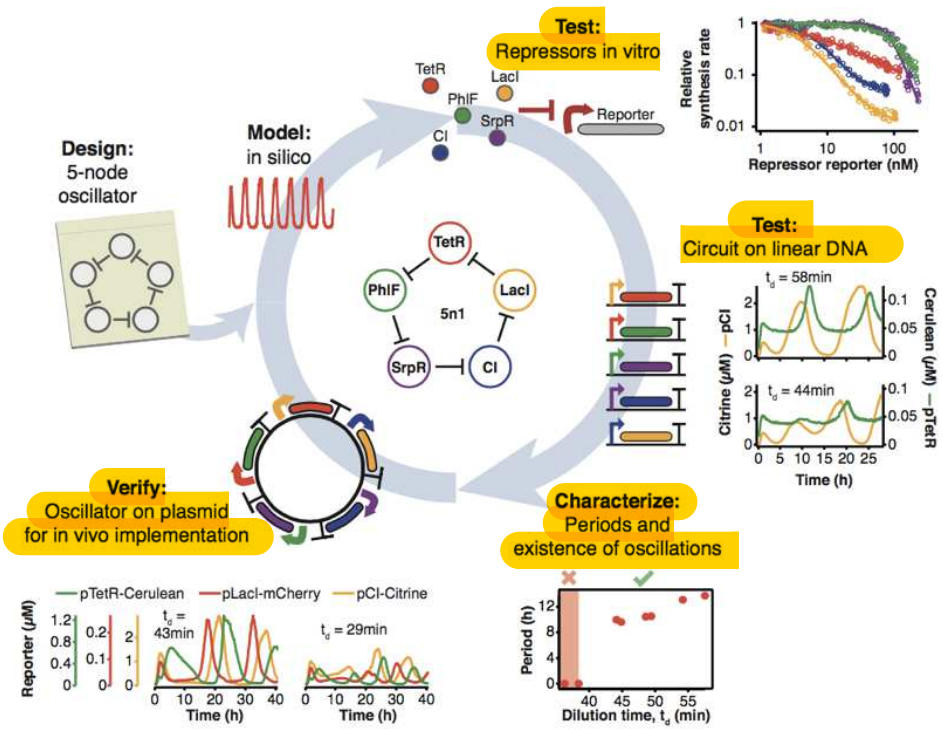


Figure 4: Synthetic biology workflow using cell-free prototyping [42].

2.2 Modeling, Analysis, and Design

In order to design the types of artificial cells and multi-cellular systems that we envision, it will be important to have an underlying mathematical framework for modeling, analysis, and design. We plan to build on concepts from (electrical) circuit theory, control theory, and stochastic systems to provide a set of tools (theory and algorithms) that can take into account stochastic dynamics of biomolecular systems, interactions between components and subsystems, and the need for multi-layered design abstractions that hide complexity of lower layer functions.

Networked systems structure A key challenge in developing models for any class of problems is the selection of an appropriate mathematical framework for the models. Among the features that we believe are important for a wide variety of biological systems is capturing the temporal response of a biomolecular system to various inputs and understanding how the underlying dynamic behavior leads to a given phenotype. The models should reflect the subsystem structure of the underlying dynamical system to allow prediction of results, but need not necessarily be mechanistically accurate at a detailed biochemical level. We are particularly interested in those problems that include a number of molecular "subsystems" that interact with each other, and so our models should support a level of modularity (with the additional advantage of allowing multiple groups to develop detailed models for each module that can be combined to form more complex models of the interacting components). Since we are likely to be building models based on high-throughput experiments, it is also key that the models capture the measurable outputs of the systems. Figure 5 shows a block diagram representation of one possible modeling framework.

For many of the systems that we are interested in, a good starting point is to use reduced-order models consisting of nonlinear differential equations, possibly with some time delay. Using the

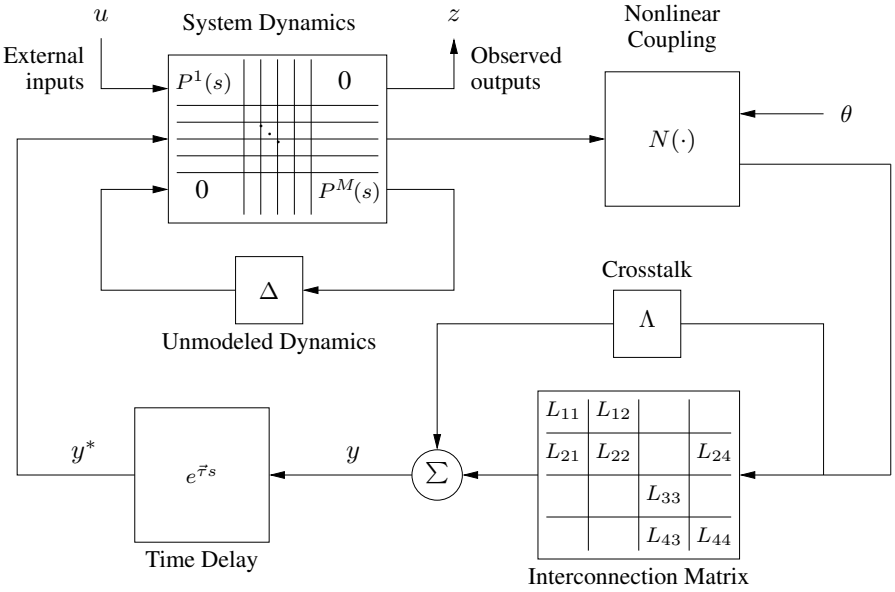


Figure 5: Interconnection analysis framework.

basic structure shown in Figure 5, a model for a multi-component system might be described using a set of input/output differential equations of the form

$$\begin{aligned} \frac{dx_i}{dt} &= Ax_i + N(x_i, Ly^*, \theta) + Bu_i + Fw_i, \\ y_i &= Cx_i + Hv_i \qquad y_i^*(t) = y_i(t - \tau_i). \end{aligned} \tag{1}$$

The internal state of the ith component (subsystem) is captured by the state $x_i \in \mathbb{R}^{n_i}$, representing the concentrations of various species and complexes as well as other internal variables required to describe the dynamics. The "outputs" of the subsystems $y_i \in \mathbb{R}^{p_i}$ describe species (or other quantities) that interact with other subsystems in the cell. The internal dynamics consist of a set of linear dynamics Ax as well as nonlinear terms $N(x, Ly^*, \theta)$ that depend both on the internal state and the outputs of other subsystems, where Ly^* represents interconnections with other subsystems and θ is a set of parameters that represent the context of the system. We also allow for the possibility of time delays (due to folding, transport or other processes) and write y_i^* for the "functional" output seen by other subsystems (in Figure 5, time delays are represented by their Laplace transform $e^{\vec{\tau}s}$).

The coupling between subsystems is captured using a weighted graph, whose elements are represented by the coefficients of the interconnection matrix L. In the simplest version of the model, we simply combine different outputs from other modules in some linear combination to obtain the "input" Ly^*. More general interconnections are possible, including allowing multiple outputs from different subsystems to interact in nonlinear ways (such as one often sees on combinatorial promoters in gene regulatory networks). The structure of L corresponds to the interactions within a subsystem (L_{ii} blocks) and between subsystems (L_{ij} blocks, where $i \neq j$).

Finally, in addition to the internal dynamics and nonlinear coupling, we separately keep track of external inputs to the subsystem (Bu), stochastic disturbances (Fw) and measurement noise (Hv). We treat the external inputs u as deterministic variables (representing inducer concentrations, nutrient levels, temperature, etc.) and the disturbances and noise w and v as random processes. If desired, the mappings from the various inputs to the states and outputs, represented by the matrices B, F and H can also depend on the system state x (resulting in additional nonlinearities).

Applicability to biomolecular systems The mathematical structure in Figure 5 and equation (1) captures a large number of modeling frameworks in a single formalism. In particular, mass action kinetics and chemical reaction networks can be represented by equating the stoichiometry matrix with the interconnection matrix L and using the nonlinear terms to capture the fluxes, with θ representing the rate constants. We can also represent typical reduced-order models for transcriptional regulatory networks by letting the nonlinear functions N represent various types of Hill functions and including the effects of mRNA/protein production, degradation and dilution through the linear dynamics. These two classes of systems can also be combined, allowing a very expressive set of dynamics that is capable of capturing many relevant phenomena of interest in molecular biology.

Another appealing feature of this formalism is that variants of it are well-studied and characterized in the control systems literature. For example, the effect of the nonlinearities can be studied using the method of harmonic balance [32] or the related technique of describing functions [4, 20]. Similarly, in the absence of the nonlinearities and with simplifying assumptions on the linear dynamics, the effect of the interconnection topology can be captured by investigating the location of the eigenvalues of the interconnection matrix (graph Laplacian) L [15].

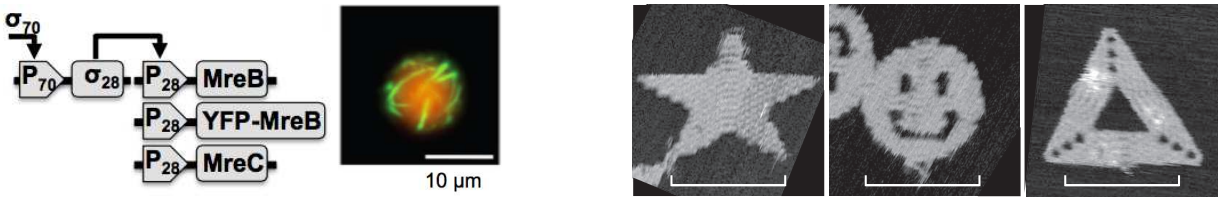
Despite being a well-studied class of systems, there are still many open questions with this framework, especially in the context of biomolecular systems. For example, a rigorous theory of the effects of crosstalk, the role of context on the nonlinear elements, and combining the effects of interconnection, uncertainty and nonlinearity is just emerging. Adding stochastic effects, either through disturbance and noise terms, initial conditions, or in a more fundamental way (see, for example, [12, Ch. 4]), is also largely unexplored, as is incorporation of crosstalk (Λ) and unmodeled dynamics (Δ). And the critical need for methods for performing model reduction in a way that respects of the structure of the subsystems has only recently begun to be explored [2, 48, 49, 54].

Computational tools and algorithms The mathematical techniques described above provide a basic framework for modeling, analysis, and design. To be useful, these techniques must be implemented as computational algorithms that can be used by circuit-, subsystem-, and system-level engineers to carry out a design at the appropriate layer of abstraction. Recent results have become available that demonstrate these types of computer aided design tools, such as NUPACK [72], Cello [43] and the GRSM compiler [52]. My group's work on biomolecular systems has so far been focused primarily on modeling [65], but we have substantial experience in other areas on the implementation of design-oriented tools [14, 37, 69]. As part of this project, we will implement our modeling, analysis, and design techniques as open source algorithms that interact with the growing body of biodesign automation tools.

Fundamental research contributions The use of model-based approaches to synthetic biology is still in its infancy. Most groups today use models as a means of explaining their experimental results, rather than as predictive tool for design. The research proposed above will provide new methods for stochastic modeling, system identification, circuit, subsystem and multi-cellular analysis, and design of cooperative behaviors. While driven by the problem of genetically-programmed artificial cells and multi-cellular machines, we anticipate that many of the techniques that we develop will be useful in other areas of synthetic and systems biology.

2.3 Circuits and Subsystems

To build an artificial cell of the sort conceptualized in Figure 1, a variety of system functions will need to be implemented. We envision that each of these functions would represent a collection



(a) Phospholipid vesicles (b) DNA origami [53]

Figure 6: Spatial localization technologies.

of biomolecular constructs (scaffolds, circuits, pathways) that operate to together as a subsystem. In this section we provide an overview of the key subsystems shown in Figure 1 and how they might be implemented in an artificial cell. As noted elsewhere, we do not propose here to develop *all* of these technologies, but rather to focus on developing and integrating a smaller subset (3–4) that would demonstrate the key concepts and serve as driver for the modeling, analysis, and design tools described in Section 2.2.

Spatial organization A key element of an artificial cell will be the ability to spatially localize biomolecular components, in the same way that natural organisms use scaffolds, organelles, cell walls, and other mechanisms. Two possible technologies that can be used to provide spatial organization are phospholipid vesicles and (3D) DNA origami. Both have already been demonstrated to some extent. Technical challenges include efficient means of encapsulating the desired genetically-programmed elements within the container and identifying means of transferring signals and molecules across the container boundaries.

Phospholipid vesicles A starting point for artificial cells is the development of phospholipid vesicles that can contain TX-TL extract, as shown in Figure 6a. In this figure (courtesy V. Noireaux), membrane-bound proteins are expressed in TX-TL in a 10 μm phospholipid droplet. This technology provides a potential means to create spatially isolated volumes that can import and export small molecules, as well as serve as a scaffold for sensing and actuation machinery. My group has used phospholipid vesicles for testing and has the capability to incorporate this technology into our workflow. Furthermore, we maintain active collaborations with the Noireaux lab, who continues to refine this technology and with whom we would plan to collaborate on this project, as needed. Recent advances in this technology have demonstrated capabilities similar to what is required for this project [1].

DNA origami Over the past decade, structural DNA nanotechnology has allowed us to create molecular programs that self-assemble into arbitrary shapes and patterns. In work at Caltech, Paul Rothemund has pioneered the technique of DNA origami to create custom patterns and shapes up to 100 nanometers per dimension in size [53], as illustrated in Figure 6b. These structures arise out of a diffusional dance in solution between 200 short "staple" strands and a single 7000-base "scaffold" strand. There are already dozens of groups around the world using DNA origami, demonstrating its robustness as a molecular self-assembly platform for diverse nanodevices.

In this project, the potential role of DNA origami is twofold: as a structural platform for implementing an artificial cell and as a scaffold for providing spatial localization of circuits. Both of these purposes have already been demonstrated to some extent: 3D DNA origami containers have been built that can isolate chemicals from their environment [13], and various DNA scaffolds have been used to localize DNA, RNA, and proteins to fixed locations [10, 55]. My group has active

collaborations with Rothemund, and my students have served as TAs for courses at Caltech that make use of DNA origami technology, thus facilitating the integration of this technology into our workflow, as needed.

Metabolism Metabolic subsystems will be responsible for providing both the energy required for the system to operate as well as small molecules, proteins, and other species required in the operation of the artificial cell. We have already demonstrated the ability to implement simple metabolic processes within TX-TL [70, 40, 73], but scaling this up to provide a (minimal) metabolic network is a major challenge. Cell extract can be used initially to provide the required functionality, with the possibility of re-energizing natural metabolic pathways [29].

An example of the sort of metabolic pathway that can be implemented in a cell-free system is shown in Figure 7. This figure shows the implementation of a pathway for production of 1,4-butanediol (BDO), which was used as part of a joint project with Genomatica, Inc [70]. This pathway is an example of the sort of small molecule metabolic pathway that is possible today. Other work in using TX-TL for metabolic pathways include demonstration of the polypeptide valinomycin [73] (done by an undergraduate in my group), 2,3-BDO [30] and violacein [40].

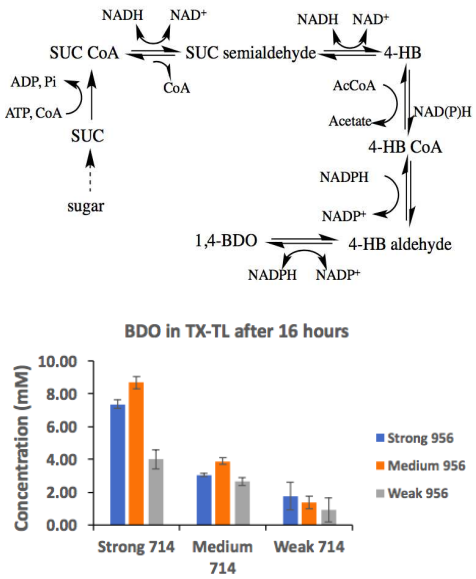


Figure 7: Cell-free synthesis of 1,4-BDO [70].

Sensing and signaling In order for artificial cells to interact with the environment, they will need to sense external conditions. Vesicle- or origami-bound proteins are natural way approaches to explore, building off of the various molecular sensors provided in natural biological systems. In addition, signaling cascades will be needed to amplify and process environmental signals.

To date, the incorporation of proteins into artificial vesicles or origami that allow signal transduction has not been demonstrated, but many groups are working on such technologies. In my own group, we have recently collaborated with Amgen, Inc. to demonstrate the ability to express membrane bound proteins in TX-TL using phospholipid "nanodisks" that emulate the cell membrane. We used these nanodisks to explore the functionality of a two-component signaling system operating in TX-TL [22], as shown in Figure 8. Specifically, we have tested $\beta 2$ adrenergic receptor (BCAR) proteins binding to an SPR surface coated with norepinephrine, with and without carazolol, demonstrating proper protein function through a fluorescence-based carazolol binding assay.

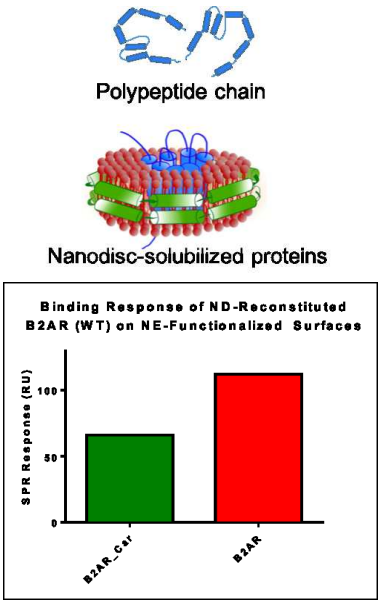
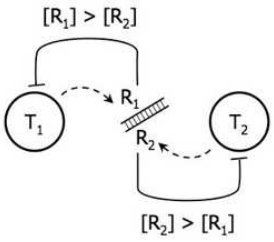
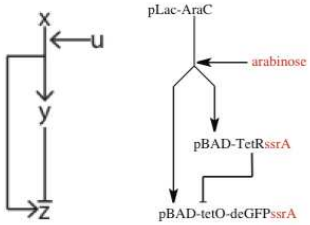


Figure 8: Expression of membrane-bound proteins using nanodisks.

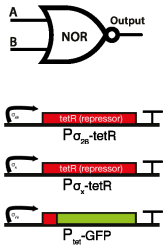
In addition, we have also demonstrated the functionality of sensing and signaling-like circuits using genelet-, transcription- and phosphorylation-based components and transcriptional regulation. Genelet components include an incoherent feedforward-loop that allows adaptation to signal levels [34] and an insulation



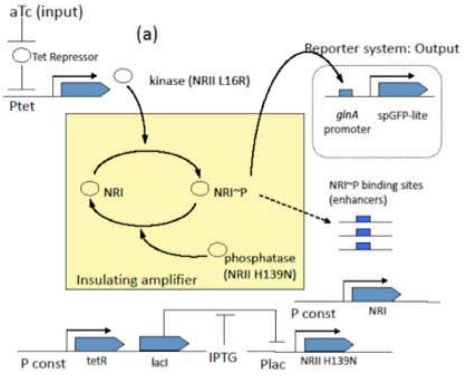
(a) Genelet rate regulator [17]



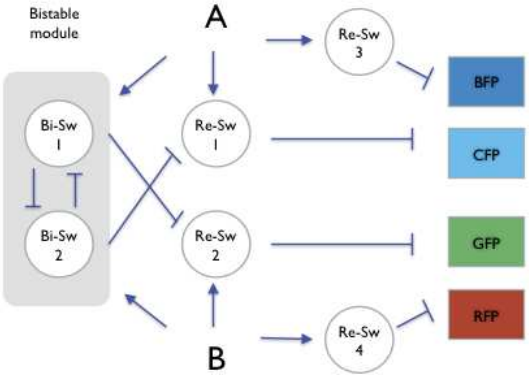
(b) Incoherent feedforward loop [21]



(c) TX-TL logic gate [25]



(d) Phosphorelay-based insulator [23]



(e) A then B event detector (unpublished)

Figure 9: Cell-free regulatory and computational circuits.

circuit that allow minimization of circuit loading [16]. Transcriptional circuits that might be used in signaling pathways include a molecular sensor to detect vanillin [11] and a fold-change detection circuit [21]. Finally, a phosphorylation-based "insulator" was tested in TX-TL and demonstrated the ability to reduce circuit loading [23]. Each of these circuits serves as a source of experience that can be utilized for implementing more complex sensing and signally subsystems.

Regulation and computation Central to the operation of the artificial cell is the ability to regulate the various operations of its subsystems and to perform computations that enable interesting behaviors and functions. This is perhaps the area that is most well-studied in my group's past research, although the complexity of circuits we are currently able to build is far below what would be required for even a minimal artificial cell.

A number of preliminary results are available demonstrating implementation of regulatory and computational elements in cell-free settings. Genelet technology has been used to create oscillators [36], oscillators driving a set of DNA "tweezers" [16], a feedback controller to regulate the the rates of production of RNA [17], and the previously-mentioned incoherent feedforward loop [34]. We have also demonstrated the implementation of a set of two input logic gates in TX-TL, as part of a demonstration project at the DARPA "Wait, What?" conference [25] (now being transitioned for use in high school laboratories as part of the BioBuilder curriculum). Figure 9 illustrates some of these circuits.

In addition to the use of cell-free circuits for regulation and computation, we also anticipate that many cell-based technologies will be amenable to implementation in artificial cells and multicellular machines. Work in my group that is directly relevant includes work with scaffold proteins

to implement a concentration tracker [27], use of integrase circuits for temporal logic gates [28], and ongoing work in the use of CRISPR-based guide RNA logic (done jointly with Niles Pierce's group at Caltech).

Actuation and locomotion In some instances, it may be useful for artificial cells to be able to move in their environment, perhaps using cilia- or flagella-like mechanisms. Another possibility is to control the shape of cells using filaments or other cytoskeletal structures. This is an area of high risk, since the molecular actuators that are present in biological systems are much more complicated than anything that has been demonstrated in cell-free environments to date.

There are multiple approaches that might be taken to move forward this particular subsystem technology. One possibility would be to make use of DNA nanostructures such as the tweezers that were used in previous joint work with Eric Winfree and Fritz Simmel [16]. A much more complicated approach would be to attempt to implement the flagellar machinery of a microorganism, perhaps first by using nanodisks for prototyping the membrane-bound machine that is found in living cells.

Materials input/export and communications Another useful function for artificial cells is export of small molecules or proteins out of the cell's container and into the environment. These can be used for communication between cells, materials production, or export of enzymes. Although there are a variety of export proteins available in nature, none have been demonstrated to work in artificial vesicles and (3D) DNA origami approaches have so far demonstrated limited capabilities of this sort.

Given our ability to prototype membrane-bound circuits in cell-free systems, we anticipate exploring the use of various chemical transporters and efflux pumps as a means of exploring this direction. We are also pursuing two independent projects in the use cell-based multi-cellular communication using AHLs (J. Parkin and R. M. Murray, 2016 [unpublished]). The use of microfluidically controlled environments and phospholipid vesicles provides a starting put for moving these cell-based technologies into an artificial cell setting.

Fundamental research contributions The concepts listed above represent an approach to developing a collection of circuit and subsystem functions in an artificial cell context. Some of them will involve transitioning concepts from other settings into an artificial cell environment. Others will undoubtedly evolve into more fundamental research projects that focus on the core mechanisms involved and provide new insight into how biomolecules function and how we can engineer novel behaviors using existing and new components. Importantly, it is the combination of the analytical approaches described in Section 2.2 with the experimental efforts in this section where we see the most potential for new fundamental research advances.

In addition to the specific circuits and subsystems, an underlying need is to develop methods for design of *programmable* genetic elements for implementing biomolecular circuits. Three leading technologies are genelet-based circuits (or other RNA-based methods, such as those developed by Julius Lucks []), integrase-based circuits, which can be programmed through arrangement of (possibly nested) attachment cites [6, 28, 58], and CRISPR-based circuits, which can be programmed through the use of guide-RNAs whose sequence and/or secondary structure can be used to modulate the dynamics of the circuit [19, 44, 45, 50]. The overriding feature that is needed is the ability to program the operation and interaction between circuits and subsystems, which will serve as an enabler for building systems with hundreds of engineered biological components (built by multiple groups, using a interoperable approach).

2.4 Systems Engineering and Technology Demonstrators

In addition to the individual subsystems and functions described above, two major challenges will be the integration of the subsystems and assembly of the overall artificial cell. It is likely that compartmentalization will be required to limit the interactions between subsystems (similar to the spatial organization present in natural cells). How to assemble the various subsystems into a functioning artificial cell is also a major challenge. One possibility would be to encapsulate the DNA encoding the various functions into a vesicle, as described above, and then establishing an "assembly" process by which the subsystems would self assemble and integrate into the chassis wall (where appropriate).

To address these issues and to demonstrate the results of the theoretical framework (Section 2.2) and circuit/subsystem designs (Section 2.3), we plan to explore the development of two or more "system-level" technology demonstrators. Each of these demonstrators will bring together multiple subsystems in a way that creates a more complex set of functions than currently available and provides insights into how more complex artificial cells and multi-cellular machines can be created. In this section we speculate on some of the possible demonstrators that we could pursue, although these are likely to change as the project unfolds and new technologies are developed (by us and by others).

Distributed event detection One of the DoD relevant applications of biologically engineered systems is the detection of small molecules (chemical signatures) and the monitoring and logging of sequences of events. An "event detector" is a circuit that allows the detection of a pattern of chemical inputs that might vary in terms of species combinations, relative magnitudes, and temporal timing. An "event logger" is a circuit that records a sequence of events (or environmental states) in a manner that can recovered at a later time. Preliminary research includes the detection of small molecules [11, 18, 59], implementation of logical functions [7, 38, 58, 67], detection of sequences of events [28, 52], and methods for implementing long-term memory [58]. My group has experience with all of these technologies (as described in Section 2.3) and the capability to integrate the works of others, as appropriate.

As a possible demonstration of the broader concept of artificial cells and multi-cellular machines, one can envision encapsulating one or more of the components of an event detection and logging circuit in individual artificial cells (perhaps phospholipid vesicles as a starting point) and using small molecules as a means for components that are in different cells to communicate with each other. This would be similar to the conceptual diagram shown in Figure 10, where the square boxes represent individual functions (detectors, logic, memory) and the small circles represent chemical signals. Preliminary demonstrations of vesicle-based distributed operations have recently been demonstrated by Adamala and Martin-Alarcon et al. [1], and our own work [28] has demonstrated the use of integrase-based circuits as a means of both implementing temporal event detection ("a then b" logic) and providing long term memory.

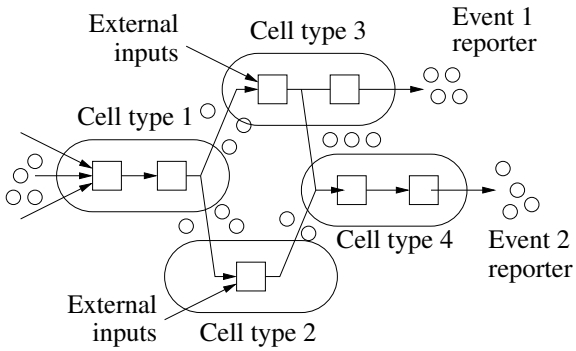


Figure 10: Distributed sensing concept.

There are several advantages of using multi-cellular event detection techniques. By creating

"modules" that implement common subfunctions, a variety of behaviors can be created by controlling which circuit elements are combined together in a multi-cellular machine. In addition, by exposing a population of artificial cells to environmental conditions and measuring the distribution of the response, it is possible to achieve new types of measurement and control functionality. For example, our recent work using integrase-based event detectors demonstrated the ability to infer timing, duration and amplitude of pulses of chemical events by looking at the distributional response of a population of cells [28].

Distributed chemical conversion A second possible demonstration is the implementation of a complex metabolic pathway (or combination of pathways) in a collection of artificial cells. The idea here would be to convert one or more chemical substrates into one or more chemical products, using individual artificial cells as a means of implementing some portion of the pathway. This approach is illustrated in Figure 11, where the square boxes now represent metabolic reactions and the small circles represent chemicals that are being converted from the substrate (leftmost triangles) to one or more products (rightmost pentagons and hexagons).

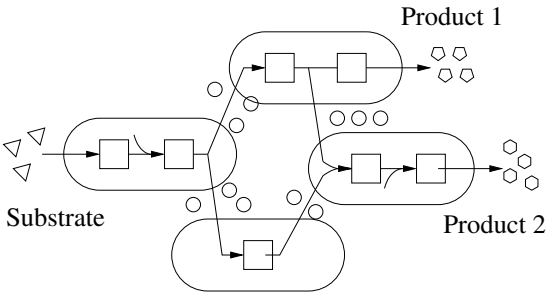


Figure 11: Distributed metabolic pathway.

While similar at a diagram level to the distributed event detection demonstration, this multicellular system would have several unique characteristics. First it would be necessary to implement metabolic pathways in the artificial cells, a capability that seems feasible given recent results on cell-free metabolic engineering [40, 70, 73], but that has not been demonstrated. Second, it would be necessary to import and export the appropriate metabolic intermediates, which might involve the implementation of efficient transporters for the molecules of interest. Third, depending on the molecules that are present in the pathway, there may be issues with "crosstalk" (import of non-target substrates into an artificial cell), availability of co-factors, and potential "toxicity" (undesired reactions that interfere with the operation of the desired conversion reactions).

Flagellar-controlled locomotion As a final example of a possible demonstration, and one that is considerably higher risk than the others, we consider the implementation of a flagellum-like actuator in an artificial cell. The flagellum represents a complex molecular machine that is fairly well understood, as illustrated by the diagram in Figure 12 [66]. Recent studies on the structure of the flagellum provide information into its molecular components [9], its mechanical operation [61] and its use as an actuator in complex behaviors such as search [5, 51].

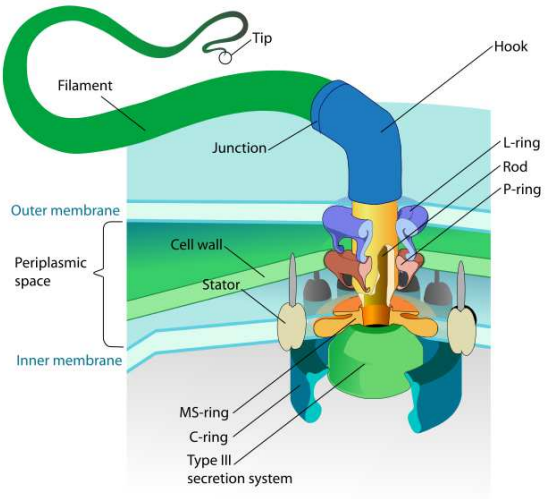


Figure 12: Gram-negative bacterial flagellum.

One can envision building up the components of a flagellum and implementing them in an artificial cell (albeit one with a more complex membrane structure). The various proteins used in the flagellar machinery could be expressed both individually and in combination to understand and replicate the self-assembly that takes place in living systems. If the components can be (self-) assembled in

an *in vitro* setting, this could provide deeper insight into the processes of assembly and operation of the flagellum, and also how its movement can be controlled and modulated.

Fundamental research contributions While the examples described above could be considered as just demonstrations of a set of technologies, proper execution of these demonstrations will serve as a basis for several different fundamental research contributions in cell-free (and likely cell-based) synthetic biology. We briefly summarize three such potential contributions here.

Multi-element ensembles for distributed computation To date, most examples in synthetic biology involve circuits that operate in single cells. While there is are many activities aimed at multi-cellular systems and consortia, there has not yet been a concerted effort to understand the system-level properties for such systems. The stochastic nature of biological systems creates a challenge, since genetically identical components have (large) distributions of responses. This highly stochastic behavior will require mathematical approaches that are likely different than those used in other engineering disciplines, where it often suffices to use Gaussian distributions and keep track of just the first two moments of a unimodal response. One potential advantage of the highly stochastic nature of biological systems in this distributed setting is the ability to use population-level response to obtain analog information from binary decisions (as illustrated in our event detector work [28]) and the ability for populations to function across a broader set of operating environments.

Managing crosstalk and uncertainty A fundamental challenge in synthetic biology is the creation of circuits that have have hundreds or thousands of components. While managing complexity is one of the limiting factors in design of such circuits, two other key aspects are crosstalk between components and uncertainty in the components themselves. Natural biological systems operate in the presence of both crosstalk and uncertainty, and in many instances exploit such features to obtain robust operation across a wide range of conditions. In other engineering disciplines, we seek to minimize crosstalk and uncertainty to enable implementation of complex behaviors. This may not be possible or desirable for biological systems. The system-level demonstrations proposed here will force us to confront these issues and to either find ways to manage crosstalk and uncertainty or, at worst, to provide additional insights and motivation for future avenues of research.

Testbeds and characterization Based on past experience, a substantial amount of activity is likely to take place in the construction of test environments for artificial cells and multi-cellular machines and the tools used to characterize their behavior. New methods for constructing, testing, and characterizing engineered biological systems have the potential to provide new insights and new capabilities that accelerate advances in modeling, design, and implementation. An example of the power of such advances is evident in our joint work with Sebastian Maerkl at EPFL, who is an expert in construction of microfluidic devices and is interested in cell-free systems. My group established a collaboration with Maerkl's group that made use of a microfluidic technology to provided the capability of continuous expression of proteins with mixing of fresh extract and DNA into a microfluidic ring [41]. This technology was a key enabler for the workflow shown in Figure 4 and is now part of our biomolecular breadboards capability (column D in Figure 3).

2.5 Dissemination and Interaction

In addition to the development of the basic technology, a major element of the proposed effort will be the dissemination of tools for engineering biology to the broader community. We will use a variety of mechanisms, including establishing collaborations with leading research groups who

are interested in the component and system technologies, visiting and interacting with DoD researchers to identify research opportunities, offering short courses to expose interested researchers to the underlying technologies, and maintaining "open source" repositories of algorithms, protocols, and techniques. These activities are already a central part of our current activities in cell-free synthetic biology, including running three TX-TL workshops, maintaining all protocols on the OpenWetWare web site, and providing support for approximately 15 groups (academic, industry, and government) who are using the biomolecular breadboards that we have helped to develop.

3 Work Plan

The following high-level milestones will be used to accomplish the overall project objectives.

Year 1 Establish the initial approach to design and implementation of circuits and subsystems in artificial cells, as well as connections to DoD applications and researchers: • Implement at least two input/output circuits or pathways in an artificial cell and characterize their performance. • Make use of the mathematical framework in Section 2.2 to model all circuits and subsystems. • Establish connections with at least three DoD labs that can serve as partners for research in synthetic biology, and run a "boot camp" on cell-free synthetic biology.

Year 2 Demonstrate the ability to design and implement a "subsystem" using prior work and results from Year 1: • Implement and characterize at least one biomolecular subsystem capable of operating in a spatially-isolated, cell-free environment (vesicle- or origami-based). • Expand the mathematical framework of to account for uncertain behavior, including unmodeled dynamics Δ and crosstalk Λ. • Host one or more DoD visitors for multi-week (ideally multi-month) visits to participate in the project, and run a second "boot camp" on cell-free synthetic biology.

Year 3 Design and implement multiple subsystems, and prototype a method of assembling them into a multi-cellular machine: • Implement three or more biomolecular subsystems operating in spatially-isolated, cell-free environments and characterize their robustness properties. • Demonstrate one or more methods for assembling multiple biomolecular subsystems into a multi-element ensemble with structured interactions between the individual subsystems. • Make use of the expanded mathematical framework from Year 2 as an integral element of the design of all subsystems. • Host multiple DoD visitors for a multi-week or longer visits and run a workshop on artificial cells.

Year 4 Optimize designs and establish robustness properties of multi-cellular machines: • Characterize and optimize the biomolecular subsystems from Year 3 to provide robust and modular operations. • Develop and demonstrate a suite of design tools (mathematical framework and supporting software) for modeling, analysis, and design of circuits, subsystems and (artificial) cells. • Optimize methods of assembling multiple biomolecular subsystems into a multi-element ensembles and disseminate the results via open source repositories. • Transition at least one circuit or subsystem to a DoD application in collaboration with a relevant DoD laboratory.

Year 5 Refine prior results, demonstrate the relevance of our research to selected applications, and identify areas for future research. • Demonstrate at least two multi-subsystem machines that validate the proposed design framework, subsystem implementations, and assembly techniques. • Document and disseminate a suite of design tools for modeling, analysis, and design of circuits, subsystems, (artificial) cells, and multi-cellular machines. • Run a workshop on cell-free synthetic biology (artificial cells and multi-cellular machines) that involves collaborators from DoD and academia, with the goal of highlighting work in the field and identifying future research challenges.

References

[1] K. P. Adamala, D. A. Martin-Alarcon, K. R. Guthrie-Honea, and E. S Boyden. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nature Chemistry*, 2016. DOI 10.1038/nchem.2644.

[2] J. Anderson, Y.-C. Chang, and A. Papachristodoulou. Model decomposition and reduction tools for large-scale networks in systems biology. *Automatica*, 47(6):1165–1174, 2011.

[3] G. Artavanis, V. Hsiao, C. A. Hayes, and R. M. Murray. The role of single occupancy effects on integrase dynamics in a cell-free system. In *Synthetic Biology: Engineering, Evolution and Design (SEED) Conference*, 2016. Available at `http://biorxiv.org/content/early/2016/06/18/059675`.

[4] D. P. Atherton. *Nonlinear Control Engineering*. Van Nostrand, New York, 1975.

[5] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[6] J. Bonnet, P. Subsoontorn, and D. Endy. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences*, 109:8884–8889, 2012.

[7] J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, and D. Endy. Amplifying genetic logic gates. *Science*, 340(6132):599–603, 2013.

[8] F. Caschera and V. Noireaux. Compartmentalization of an all-E. coli cell-free expression system for the construction of a minimal cell. *Artificial Life*, 22(2):185–195, 2016.

[9] S. Chen, M. Beeby, G. E. Murphy, J. R Leadbetter, D. R. Hendrixson, A. Briegel, Z. Li, J. Shi, E. I. Tocheva, A. Müller, Megan J. D., and G. J. Jensen. Structural diversity of bacterial flagellar motors. *The EMBO Journal*, 30(14):2972–2981, 2011.

[10] Y.-J. Chen, B. Groves, R. A. Muscat, and G. Seelig. DNA nanotechnology from the test tube to the cell. *Nature Nanotechnology*, 10(9):748–760, 2015.

[11] E. L. C. de los Santos, J. T. Meyerowitz, S. L. Mayo, and R. M. Murray. Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: Developing a vanillin sensor. *ACS Synthetic Biology*, 2015. DOI: 10.1021/acssynbio.5b00090.

[12] D. Del Vecchio and R. M. Murray. *Biomolecular Feedback Systems*. Princeton University Press, 2014.

[13] S. M. Douglas, I. Bachelet, and G. M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834, 2012.

[14] R. M. Murray et al. Python-control: Control systems library for Python. `http://github.com/python-control`.

[15] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(5):1465–1476, 2004.

[16] E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, and Friedrich C. Simmel. Timing molecular motion and production with a synthetic transcriptional clock. *Proceedings of the National Academy of Sciences*, 108(40):E784–E793, 2011.

[17] E. Franco, G. Giordano, P.-O. Forsberg, and R. M. Murray. Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synthetic Biology*, 3(8):589–599, 2014.

[18] C. E. French, K. de Mora, N. Joshi an A. Elfick, J. Haseloff, and J. Ajioka. *Institute of Medicine (US) Forum on Microbial Threats. The Science and Applications of Synthetic and Systems Biology: Workshop Summary*, chapter A5. Synthetic Biology and the Art of Biosensor Design. National Academies Press, 2011.

[19] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S Qi. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–451, 2013.

[20] D. Graham and D. McRuer. *Analysis of Nonlinear Control Systems*. Wiley, New York, 1961.

[21] S. Guo, Y. Hori, and R. M. Murray. Systematic design and implementation of a novel synthetic fold-change detector biocircuit in vivo. In *Winter q-bio Conference*, 2015. `http://www.cds.caltech.edu/~murray/preprints/ghm15-wqbio_s.pdf`.

[22] S. Guo, A. Vaish, Q. Chen, and R. M. Murray. Expressing biologically active membrane protein in a cell-free transcription-translation platform. Technical report, bioRxiv, 2017. Submission pending approval from Amgen.

[23] S. Guo, E. Yeung, K. Soli Nilgiriwala, D. Del Vecchio, and R. M. Murray. Implementation and simulation of phosphorylation-based insulator in transcription-translation platform. Technical report, California Institute of Technology, 2014. `http://www.cds.caltech.edu/~murray/preprints/guo+14-wqbio_s.pdf`.

[24] C. Hayes and N. Kuldell. Cell-free circuits: Molecular diagnostics outside the cell. BioBuilder Laboratory Notes, 2017. In preparation.

[25] C. Hayes and R. M. Murray. DARPA "Wait, What?": Build a DNA circuit. Web site, September 2015. `http://txtlbio.org/waitwhat`.

[26] C. E. Hodgman and M. C. Jewett. Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering*, 14(3):261–269, 2012.

[27] V. Hsiao, E. L. C. de los Santos, W. R. Whitaker, J. E. Dueber, and R. M. Murray. Design and implementation of a synthetic biomolecular concentration tracker. *ACS Synthetic Biology*, 2014. DOI 10.1021/sb500024b.

[28] V. Hsiao, Y. Hori, P. W. K. Rothemund, and R. M. Murray. A population-based temporal logic gate for timing and recording chemical events. *Molecular Systems Biology*, 12(5):869, 2016.

[29] M. C. Jewett and J. R. Swartz. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1):19–26, 2004.

[30] J. E. Kay and M. C. Jewett. Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol. *Metabolic Engineering*, 32:133–142, 2015.

[31] A. S. Khalil and J. J. Collins. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367, 2010.

[32] H. K. Khalil. *Nonlinear Systems*. Macmillan, 1992.

[33] J. Kim, J.J. Hopfield, and E. Winfree. Neural network computation by in vitro transcriptional circuits. *Advances in Neural Information Processing Systems (NIPS)*, 17:681–688, 2004.

[34] J. Kim, I. Khetarpal, S. Sen, and R. M. Murray. Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Research*, 42(9):6078, 2014.

[35] J. Kim, K. S. White, and E. Winfree. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology*, 2:68, 2006.

[36] J. Kim and E. Winfree. Synthetic in vitro transcriptional oscillators. *Molecular Systems Biology*, 7:465, 2011.

[37] M. Milam and R. M. Murray et al. NTG: Nonlinear Trajectory Generation library. `http://github.com/murrayrm/ntg`.

[38] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature News*, 491(7423):249–253, 2012.

[39] N. Nandagopal and M. B. Elowitz. Synthetic biology: Integrated gene circuits. *Science*, 333(6047):1244–1248, 2011.

[40] P. H. B. Nguyen, Y. Y. Wu, S. Guo, and R. M. Murray. Design space exploration of the violacein pathway in *Escherichia coli* based transcription translation cell-free system (TX-TL). Technical report, bioRxiv, 2015. DOI 10.1101/027656.

[41] H Niederholtmeyer, V Stepanova, and S. Maerkl. Implementation of cell-free biological networks at steady state. *Proceedings of the National Academy of Sciences*, 110(40):15985–15990, 2013.

[42] H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl. Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife*, 4:e09771, 2015.

[43] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt. Genetic circuit design automation. *Science*, 352(6281), 2016.

[44] A. A. K. Nielsen and C. A. Voigt. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology*, 10(11):763–763, 2014.

[45] L. Nissim, S. D. Perli, A. Fridkin, P. Perez-Pinera, and T. K. Lu. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell*, 54(4):698–710, 2014.

[46] V. Noireaux, Y. T. Maeda, , and A. Libchaber. Development of an artificial cell, from self-organization to computation and self-reproduction. *Proceedings of the National Academy of Sciences*, pages 3473–3480, 2011.

[47] K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. DaleyKeyser, P. Yin, and J. J. Collins. Paper-based synthetic gene networks. *Cell*, 159(4):940–954, 2014.

[48] T. P. Prescott and A. Papachristodoulou. Guaranteed error bounds for structured complexity reduction of biochemical networks. *Journal of Theoretical Biology*, 304:172–182, 2012.

[49] T. P Prescott and A. Papachristodoulou. Layered decomposition for the model order reduction of timescale separated biochemical reaction networks. *Journal of Theoretical Biology*, 356:113–122, 2014.

[50] L. S Qi, M. H. L., L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, 2013.

[51] C. V. Rao, J. R. Kirby, and A. P. Arkin. Design and diversity in bacterial chemotaxis: A comparative study in *Escherichia coli* and *Bacillus subtilis*. *PLoS Biology*, 2(2):239–252, 2004.

[52] N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu. Synthetic recombinase-based state machines in living cells. *Science*, 353(6297):aad8559, 2016.

[53] P.W.K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[54] H. Sandberg and R. M. Murray. Model reduction of interconnected linear systems. *Optimal Control Applications and Methods*, 30(3):225–245, 2009.

[55] A. Shaw, V. Lundin, E. Petrova, F. Fördös, E. Benson, A. Al-Amin, A. Herland, A. Blokzijl, B. Høgberg, , and A. I. Teixeira. Spatial control of membrane receptor function using ligand nanocalipers. *Nature Methods*, 11(8):841–846, 2014.

[56] J. Shin and V. Noireaux. An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and synthetic cell. *ACS Synthetic Biology*, 1(1):29–41, 2011.

[57] D. Siegal-Gaskins, Z. A. Tuza, J. Kim, V. Noireaux, and R. M. Murray. Resource usage and gene circuit performance characterization in a cell-free 'breadboard'. *ACS Synthetic Biology*, 2014.

[58] P. Siuti, J. Yazbek, and T. K. Lu. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, 31:448–452, 2013.

[59] S. Slomovic and J. J. Collins. DNA sense-and-respond protein modules for mammalian cells. *Nature Methods*, 12(11):1085–1090, 2015.

[60] M. T. Smith, K. M. Wilding, J. M. Hunt, A. M. Bennett, , and B. C. Bundy. The emerging age of cell-free synthetic biology. *FEBS Letters*, 588(17):2755–2761, 2014.

[61] Y. Sowa and R. M. Berry. Bacterial flagellar motor. *Quarterly Reviews of Biophysics*, 41(2):103–132, 2008.

[62] Z. Z. Sun, C. A. Hayes, J. Shin, R. M. Caschera, F.and Murray, and V. Noireaux. Protocols for implementing an *escherichia coli* based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments (JoVE)*, (e50762), 2013.

[63] Z. Z Sun, E. Yeung, C. A Hayes, V. Noireaux, and R. M Murray. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synthetic Biology*, 2013.

[64] M. K. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal, K. J. Spring, S. Al-Khabouri, C. P. Fall, V. Noireaux, R. M. Murray, and J. B. Lucks. Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synthetic Biology*, 2014.

[65] Z. A. Tuza, V. Singhal, J. Kim, and R. M. Murray. An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular 'breadboard' system. In *Proc. IEEE Control and Decision Conference*, 2013.

[66] M. Ruiz Villarreal. A gram-negative bacterial flagellum. Wikimedia Commons. `http://commons.wikimedia.org/wiki/File:Flagellum_base_diagram_en.svg`.

[67] B. Wang, R. I. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2:508, 2011.

[68] W. R. Whitaker, S. A. Davis, A. P. Arkin, and J. E. Dueber. Engineering robust control of two-component system phosphotransfer using modular scaffolds. *Proceedings of the National Academy of Sciences*, 109(44):18090–18095, 2012.

[69] T. Wongpiromsarn, U. Topcu, N. Ozay, H. Xu, and R. M. Murray. Tulip: A software toolbox for receding horizon temporal logic planning. In *Hybrid Systems: Computation and Control*, 2011.

[70] Y. Y. Wu, S. Culler, J. Khandurina, S. Van Dien, and R. M. Murray. Prototyping 1,4-butanediol biosynthesis pathway in a cell-free transcription-translation system. In *Synthetic Biology: Engineering, Evolution, and Design (SEED)*, 2015. DOI 10.1101/017814.

[71] E. Yeung, . J. Dy, K. B. Martin, A. H. Ng, D. Del Vecchio, J. L. Beck, J. J. Collins, and R. M. Murray. The effect of compositional context in synthetic gene networks: Biophysical constraints arising from relative gene orientation. *Cell Systems*, 2017. To appear. Preprint available: `http://biorxiv.org/content/early/2016/11/30/083329`.

[72] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.*, 32(1):170–173, 2011.

[73] T. Zhou. Prototyping a valinomycin biosynthesis pathway within a cell-free transcription-translation (TX-TL) system. Technical report, bioRxiv, 2016. `http://biorxiv.org/content/early/2016/12/04/091520`.

Management Approach

Core team The proposed project will be carried out by a combination of 4–6 graduate students, a full-time research technician, and up to 6 undergraduate researchers. We are requesting support on this grant for 3 full-time graduate students, with the expectation that additional graduate students supported by external fellowships (including NSF and NDSEG fellowships) would also contribute to the project. The use of a full-time research technician, combined with the ability to make use of the UW BIOFAB facility (described below), will enable the graduate students to focus their activities on modeling, design, characterization, and debugging of circuits and subsystems, along with the requisite analysis and theory. Undergraduate researchers will participate by exploring new ideas as part of summer projects (through the Caltech Summer Undergraduate Research Fellowship [SURF] program) as well as aiding in circuit/subsystem implementation, characterization, and testing.

All of the students working on this project will obtain broad experience with many aspects of biological engineering, and will able to work in many areas outside of cell-free systems (including more traditional approaches in synthetic biology). In addition, through laboratory visits and Caltech-hosted "boot camps", we will build a deeper understanding within DoD laboratories of the possibilities of synthetic biology (both cell-based and cell-free).

We also anticipate the participation of one or more visiting scientists (graduate students, postdocs, faculty and other researchers), who would participate in the project with support from their home institutions. My lab typically hosts 1–3 visiting scientists at any given time and this project would provide excellent opportunities for visitors interested in this topic. Of particular interest would be the participation of DoD Laboratory University Collaboration Initiative (LUCI) program participants.

Collaborations As indicated in the description of the research effort, there are many collaboration opportunities with researchers at Caltech and at other institutions. I plan to encourage interaction with other groups as a means of increasing the amount of activity and incorporating research advances in different fields. Planned interactions include:

- Molecular programming: Caltech has several faculty who are working in the area of "molecular programming", focused on device and circuit level technologies that are highly relevant to the circuit and systems level work proposed here. These include Niles Pierce, Lulu Qian, Paul Rothemund, and Erik Winfree, all of whom have past or current joint students and projects with my group.

- Biomolecular circuit design: There are several groups working on design of biomolecular circuits (cell-based and cell-free) with whom we have ongoing collaborations that can assist in the work proposed here. Specific researchers who have recent or current collaborations include Domitilla Del Vecchio (MIT), Eric Klavins (U. Washington), and Julius Lucks (Northwestern).

- Biomolecular subsystems: The more complex behaviors that we are focused on in this proposal are also of interest to a number of recent and current collaborators, including John Dueber (UC Berkeley; metabolic pathways), Michael Elowitz (Caltech; biological circuits, intra- and inter-cellular signaling), Michael Jewett (cell-free systems and metabolic engineering) and Vincent Noireaux (U. Minnesota; cell-free systems [TX-TL] and vesicle-based TX-TL).

• Theory: There is a growing community of researchers interested in the theory of biomolecular circuit design. My group has active collaborations with Shuki Bruck (Caltech), Domitilla Del Vecchio (MIT), John Doyle (Caltech), Antonis Papachristodoulou (Oxford), Eduardo Sontag (Rutgers), and a variety of other groups interested in the use of control theory tools for systems and synthetic biology.

Project coordination To coordinate the various aspects of this project, I anticipate having monthly project meetings involving the core team (including visitors working on the project). These monthly project meetings will be used for researchers to present their recent progress and discuss the integration of the various aspects of the project. They will be particularly important for linking together the theory and experimental work. Collaborators and their students and postdocs will be invited to participate in all project meetings. These project meetings will supplement the usual 1-on-1 meetings with students as well as my weekly group meetings. My experience with past projects of this size (including recent DARPA projects) is that focused project meetings can be very useful for a team working on a challenging research problem.

In addition to in person meetings, we will make use of active collaboration tools, including Slack (messaging), Confluence (wiki) and GitHub (open source code repository), to both enable communications and documentation of work across the team and inclusion of other project participants in our work. These tools also allow a good access control policies, including public dissemination at the appropriate stages of development. All active collaborators will have access to Slack, Confluence and GitHub resources. In addition, all genetic constructs produced in this project will be sent to Addgene for open distribution to researchers and all publications will be placed on bioRxiv (preprint repository).

Fit to the VBFF Program The research proposed here does not directly align with the "engineering biology" subtopics listed in the program solicitation. While I am interested in these areas, there is already funding available for work on these topics as part of ongoing Army and DARPA programs (of which I am part). Hence, this proposal would be focused on a "moonshot" that would develop substantial new capability beyond what is currently available and funded. Many of the advances in this proposal will benefit some of the listed areas, since establishing the components and architecture for artificial cells will require advances in devices, modeling, analysis, and design that are applicable across many areas of synthetic biology. But cell-free systems offer many intrinsic advantages, especially for the DoD, including portability (e.g., paper-based cell-free circuits), safety (via true orthogonality and lack of self replication), and stability (due to lack of mutation and evolution).

Principal Investigator (PI) Time

If funded, this project would support 20–30% of the activities currently being undertaken in my group. As such, it would represent the largest single project in my group and I anticipate that it would receive a corresponding amount of my time and attention.

Supervision of research I will be responsible for providing overall project leadership as well as supervising the research activities of the graduate students, postdocs, undergraduates, and research staff working on this project. In addition to weekly group meetings and regular 1-on-1 meetings with individual researchers, a monthly project meeting will be held with the core team as well as interested collaborators. I will also be responsible for dissemination of the research by maintaining

information repositories (Slack, Confluence, GitHub) that contain information about the project, giving talks and helping write papers that describe the research achievements, and integrating new knowledge from the project into courses and textbooks.

Participation in DoD activities I am already an active participant in DoD activities, including past participation in DARPA-sponsored Defense Sciences Study Group (DSSG) and Information Science and Technology (ISAT) study group, the Air Force Scientific Advisory Board (AFSAB), and current membership on the Defense Innovation Board (DIB). In addition, through the Army's Institute for Collaborative Biotechnology (ICB) I have had recent interactions with synthetic biology researchers at the Army Research Laboratory (ARL), the Edgewood Chemical and Biological Center (ECBC), and the Environmental Laboratory of the Army's Engineer Research and Development Center (ERDC) in Vicksburg, Mississippi. I anticipate continuing these interactions as well as establishing new relationships if selected as a Vannevar Bush Faculty Fellow (a budget of 6 domestic trips per year as been requested to support the associated travel costs).

Other obligations I have a number of federally-supported grants in the area of synthetic biology that complement the work proposed here:

• Robust Multi-Layer Control Systems for Cooperative Cellular Behaviors (DARPA Biological Control Program). The goal of this project is to develop and demonstrate a (cell-based) multi-layer intra- and inter-cellular control systems integrated to create complex, spatially-organized, multi-functional model system for wound healing. Our system makes use of a layered control architecture with feedback at the DNA, RNA, protein, cellular and population levels to provide programmed phenotypic differentiation and interconnection between multiple cell types. This project is an active collaboration with John Doyle, Michael Elowitz and Niles Pierce. My groups activities are focused on the design, testing and implementation of feedback controllers in *E. coli* to modulate growth rate in response to spatio-temporal inputs. Approximately 10-20% of my time is committed to this large project, which will continue through 2020.

• Rapid, Reliable and Repeatable Platforms for Cell-Free Prototyping (Synvitrobio SBIR). We are working with Synvitrobio, Inc. (a startup company in the San Francisco area, which I co-founded with a former PhD student) to develop and commercialize a cell-free platform for rapid expression and characterization or proteins and pathways. Less than 2% of my time is committed to this small project.

• AFOSR (BRI): Theory-based Engineering of Biomolecular Circuits in Living Cells. This is a joint AFOSR BRI project between MIT (Del Vecchio and Collins), Caltech (Murray) and Rutgers (Sontag). The objective of this research is to establish a data-driven theoretical framework based on mathematics to enable the robust design of interacting biomolecular circuits in living cells that perform complex decision making. Approximately 2% of my time is committed to this project, which ends in February 2019. There is some overlap of the topics in this proposal and those described in Section 2.2, but the AFOSR project is focused on cell-based circuits.

• ARO (Institute for Collaborative Biotechnology UARC): Biomolecular Feedback Circuits for Modular, Robust and Rapid Response. The Institute for Collaborative Biotechnology (ICB) is a joint activity between UC Santa Barbara, Caltech and MIT, with individual projects that are renewed on an annual basis. Our current project is focused on the development of a a set of integrase-based circuits for detecting and remembering complex temporal patterns, and triggering

a response. This project takes approximately 5% of my time and has some overlap with the ideas described in Section 2.4. If awarded a VBFF grant, I will likely redirect my ICB activities to other areas of interest that complement the work described in this proposal.

• ONR MURI: Next-generation Genetic Devices: Model-guided Discovery and Optimization of Navy-relevant Cell-based Sensors. This is a MURI project led by Chris Voigt at MIT and involving Domitilla Del Vecchio (MIT), Michael Laub (MIT), Vincent Noireaux (UMN), Eduardo Sontag (Rutgers), Howard Salis (Penn State), and Jeff Tabor (Rice). We are applying tools from synthetic biology to construct high-performance and robust sensors that respond to non-natural signals. My group's work on this grant focuses on the use of cell-free prototyping for signal detection circuits, with cell-based implementations as the final goal. This project takes approximately 2% of my time and will be ending in late 2018.

• Molecular Programming Architectures, Abstractions, Algorithms, and Applications (NSF). This is a joint project between Caltech (Bruck, Pierce, Qian, Rothemund, Winfree), Harvard (Shih, Yin), U. Washington (Klavins, Seelig) and UCSF (Douglas). Molecular programming involves the specification of structures, circuits, and behaviors both within living and non-living systemssystems in which computing and decision-making will carried out by chemical processes themselves. This project is aligned with the goals of this VBFF proposal and has served as the source of support for some of the preliminary work. This project takes approximately 2% of my time and will be ending in early 2018.

In addition, I am active in the area of network control systems and has two ongoing federal grants in that area:

• VeHICaL: Verified Human Interfaces, Control, and Learning for Semi-Autonomous Systems (NSF). This is a new NSF project pursuing research in semi-autonomous vehicles (cars and aircraft) that will continue through 2021. It takes approximately 2% of my time.

• DARPA/SRC (StarNet): The TerraSwarm Research Center. This is a DARPA and Industry sponsored project related to Internet of Things. This project takes approximately 5% of my time and will be ending in Fall 2017.

Many of these obligations ramp down in the 2017-18 time frame and support through this program would allow me to focus my attention on this project without having to write as many individual grants to support his activities.

Facilities

PI laboratory I have a fully equipped molecular biology laboratory with all of the equipment required to carry out the proposed research. Specific equipment available includes: fluorescent and standard light microscopes (Olympus IX81, MV10), ONIX microfluidic system (CellASIC), spectrophotometer (Beckman Coulter), LabCyte Echo acoustic liquid handler, Hamilton STARlet liquid handler, 3 Biotek H1M monochromator-based plate readers, spectrofluorimeter (Horiba) multi-image light cabinet, and real-time PCR systems (Applied Biosystem, Eppendorf). We also maintain a cell-free extract preparation facility with shakers, centrifuges, and a lyophylizer for production of cell-free extracts in frozen and freeze-dried form (including paper-based).

Caltech facilities In addition to individual equipment available in my laboratory, Caltech maintains a set of core facilities that provide additional capabilities. Relevant Caltech core facilities and services include: flow cytometry, protein purification, mass spectroscopy (including GC-MS and LC-MS), and DNA sequencing. The Kavli Nanoscience Institute (KNI) provides cleanroom facilities, including all equipment required for fabrication of microfluidic devices.

U. Washington BIOFAB The UW Biofabrication Center is a unique service laboratory operating through the Electrical Engineering Department at the University of Washington. Researchers use the Aquarium software package to store sample information and remotely submit molecular biology and microbiology jobs. Technicians in the BIOFAB execute the protocols and upload the results to Aquarium. Therefore, cloning projects can be executed from start to finish, without the user having to carry out time-consuming (and often error-prone) experiments. Records of all jobs and samples are permanently stored in the Aquarium inventory database, eliminating the need for messy lab notebooks and complex inventory systems.

Caltech has an established relationship with the UW BIOFAB and can make use of its capabilities to maximize the productivity of researchers working on this project. Current and planned UW BIOFAB capabilities include: Golden Gate assembly using MoClo components (including the CIDAR MoClo library), media and reagent preparation, TX-TL reactions (in development, jointly with Caltech), and plate reader assays (planned).

Computational resources For simulation and computational analysis, locally available computing includes two 8-core Linux computer servers with up to 64GB of memory, as well as access to a 60 node high performance computing cluster maintained by Caltech's Information Management Systems and Services (IMSS). We anticipate that most computation will be carried out on cloud-based resources such as Amazon Web Services EC2 cluster, which provides instance types with up to 128 cores and 2TB of memory. We are currently using EC2 for our system identification work.

Special Test Equipment

No special test equipment is required.

Equipment

No additional equipment is required.