Dynamic control in a coordinated multi-cellular maze solving system

Allen Hsu[‡*], Vikram Vijayan[‡*], Lawrence Fomundam[†], Yoram Gerchman[‡],
Subhayu Basu[‡], David Karig[‡], Sara Hooshangi[‡], Ron Weiss[‡§]

Abstract— **Control system theory provides convenient tools and concepts for describing and analyzing complex cell functions. In this paper we demonstrate the use of control theory to forward-engineer a complex synthetic gene network constructed from several modular components. Specifically, we present the design and simulation of a synthetic multi-cellular maze-solving system. Here, bacterial cells are programmed to use artificial cell-to-cell communication and regulatory feedback in order to illuminate the correct path in a user-defined maze of cells arranged on a surface. Simulations were used to analyze the system's spatiotemporal dynamics and sensitivity to various kinetic parameters. Experiments with *Escherichia coli* were carried out to characterize the diffusion properties of artificial cell-to-cell communication based on bacterial quorum sensing systems. The rational design process and simulation tools employed in this study provide an example for future engineering of complex synthetic gene networks comprising multiple control system motifs.**

I. INTRODUCTION

LIVING cells rely on control systems and motifs, such as bi-stable switches and oscillators, to regulate intracellular processes and produce robust responses to external stimuli. Control systems are prevalent in many cell functions, including the bacterial lactose utilization network [1], the adaptation of tumbling probabilities in bacterial chemotaxis [2], and the cell-cycle oscillations governing cell division [3]. The field of synthetic biology studies such naturally occurring processes in order to establish guidelines for the forward-engineering of artificial systems [4]. The implementation of these systems typically involves an iterative process of design, modeling, and testing.

Several synthetic prototype circuits have been created *in-vivo* using this iterative process, including the toggle switch [5], repressilator [6], and pulse generator [7]. Each circuit relies on a particular motif for operation. The toggle switch

achieves bi-stability through mutual inhibition of transcriptional regulators. The repressilator employs three transcription factors arranged in a ring to create oscillations in gene expression. The pulse generator utilizes a delay established by feed-forward regulation for transient expression of a reporter protein.

In this paper, we present the design, simulation, and component testing of a dynamic multi-cellular system that programs *E. coli* cells to solve mazes. The system was modeled and simulated as an interconnection of modules such that techniques from control theory could be applied to analyze the system. Sensitivity analysis of system components was undertaken to determine the kinetic constants crucial to system operation. Experiments have been conducted in order to characterize the artificial cell-to-cell communication. The results from experimentation were applied to the model in order to gain a better understanding of the system's dynamics and kinetic parameters, while results from simulation will help guide future experiments.

The synthetic gene networks used for programming the cells incorporate several important motifs including feed-forward regulation, positive and negative feedback, and mutual inhibition. Some of these motifs have been demonstrated recently in separate experimental systems [5] [7] [8], but a system containing all of these motifs has yet to be analyzed and tested. The realization of a maze solving system will serve as an important step toward implementing other complex multi-cellular systems that rely on multiple control systems and motifs. These systems will be useful for a wide range of biomedical applications such as pattern formation for tissue engineering or biomaterial fabrication.

II. SYSTEM DESIGN

A. Algorithm

A recursive algorithm was designed to find a path between two predefined points in a user-generated maze. Colonies are arranged in the shape of a maze, surrounded by a wall of cells. The colonies along the correct path are programmed to illuminate the solution to the maze as a

Fig. 1. Maze setup and example interactions. Dotted dark arrows represent exogenous *Signal 1,* dark solid arrows represent *Signal 1* generated by cells, and light arrows represent *Signal 2*. (a) Setup of maze. The maze is contained in a Petri dish with *Maze* and *End* colonies surrounded by a lawn of *Wall* cells. The *End* colony is placed at the end of the maze. An exogenous pulse of *Signal 1* is administered at the entry of the maze. (b) Illustration of maze solving algorithm. Each panel shows the state of the maze during a different step in the algorithm. Time = 1 shows the state of the maze during the first exogenous pulse of *Signal 1*, and Time = 9 shows the maze solving at its completion.

result of executing the algorithm. Fig. 1a shows the composition of a particular maze that can be solved by the algorithm. The system is composed of two types of cells denoted *End/Maze* and *Wall*. The user induces *End/Maze* cells to enter either the *Maze* or *End* state before placement of the colonies. *Maze* colonies are then placed in the shape of the maze and one *End* colony is positioned at the end as demonstrated in Fig 1a. *End/Maze* cells communicate using two diffusible cell-to-cell communication signals denoted *Signal 1* and *Signal 2*. *Wall* cells define the boundaries of the maze by preventing the communication between non-adjacent *End* and *Maze* colonies.

To initiate the maze solving algorithm, exogenous *Signal 1* is placed at the entry of the maze. In response, the first *Maze* colony emits a pulse of *Signal 1,* which diffuses to neighboring colonies. The second *Maze* colony then responds to *Signal 1* by secreting a pulse of *Signal 1*. The process continues until *Signal 1* reaches the *End* colony which secretes *Signal 2* in response. *Signal 2* diffuses to all neighboring colonies. *Maze* colonies that encounter *Signal 2* transform into *End* colonies, while *End* colonies encountering *Signal 2* remain *End* colonies. Once this transformation takes place, and sufficient degradation of

Fig. 2. (a) Mealy Finite State Machine (FSM) representation of the maze solving system. The finite state diagram shows the output as a function of current state and inputs. A slash separates the input from the resulting output action. (b) Communication between neighboring cells in the *Maze* and *End* state. PPG stands for propagating pulse generator, TS for toggle switch, SR for signal router, and SS for signal sensor. The TS can be set (S) to the *End* state (E), or reset (R) to the *Maze* state (M). Components inactive in a given state are shaded gray, while inactive pathways are dotted. Initially, the two cells on the left are in the *Maze* state and the cell on the right is in the *End* State. Exogenous *Signal 1* is fed into the PPG (1), which pulses *Signal 1* in response. *Signal 1* diffuses into the neighboring *Maze* cell (2), which propagates *Signal 1* to the neighboring *End* cell (3). The SR in the *End* cell routes *Signal 2* in response, which diffuses into the neighboring *Maze* cell (4). *Signal 2* is sensed by the SS which sets the TS in the cell to the *End* state (5), making both cells on the right *End* cells.

the communication signals has occurred, exogenous *Signal 1* can be re-administered to start another iteration of the algorithm. When the number of iterations equals the number of colonies between the entry and end of the maze, the maze solving is complete, and the path from start to finish is indicated by the positions of *End* colonies. *End* colonies are designed to fluoresce such that the path from start to finish will be illuminated. It is important to note that the system is not perfect and might lead to slight inaccuracies at junctions. A single *End* colony may be established on the path leading to the dead end, but with large enough mazes, these inaccuracies are negligible. Fig. 1b illustrates the various steps involved in reaching the solution.

B. High Level Description

Fig. 2a shows the maze solving system using a Mealy Finite State Machine (FSM) representation [9]. The diagram highlights the behaviors necessary for the implementation of the algorithm described in the previous section. This representation was decomposed into several smaller control systems that are interconnected as shown in Fig. 2b. Each modular component was designed and

simulated independently before incorporation into the complete system. The components include a propagating pulse generator, toggle switch, signal router, signal sensor, and signal degrader.

C. Circuit Components

The abstract high level description mentioned in the previous section was converted into concrete biological parts and circuits. This section explains the biological parts in each of the components designed for the maze solving system. Fig. 3 depicts the interconnections between the genetic circuits in the *End/Maze* cell.

1) Signals: Acyl-homoserine lactones (AHLs), a family of bacterial quorum sensing signals, were chosen to implement the two communication signals. AHL molecules can be synthesized by 'I' proteins, are small enough to diffuse in and out of the cell membrane, and can activate transcription when bound to their cognate 'R' protein [10]. Appropriate communication signals have sufficient sensitivity and minimal crosstalk between them. Based on reports in literature and preliminary experimental results in our lab, *Signal 1* and *Signal 2* were chosen to be 3-oxododecanoyl-homoserine lactone (3OC12HSL) and butanoyl-homoserine lactone (C4HSL) respectively [11].

2) Propagating Pulse Generator (PPG): The propagating pulse generator synthesizes 3OC12HSL transiently in response to 3OC12HSL. The PPG circuit shown in Fig. 3b is composed of two network motifs: a feed-forward loop and negative feedback. Presence of 3OC12HSL triggers production of additional 3OC12HSL and CI. CI represses further production of 3OC12HSL, resulting in a temporary pulse. The PPG is shut down after a pulse due to high concentrations of CI, and only resets once both 3OC12HSL and CI degrade. The circuit extends the design of the pulse generator [7] with an additional intercellular feedback of LasI to generate a signal propagating pulse.

3) Toggle Switch (TS): The toggle switch shown in Fig. 3a is responsible for maintaining the state of the *End/Maze* cell. The TS can enter either the *Maze* or *End* state in response to external stimuli. The *Maze* and *End* states have their own corresponding fluorescence reporters which are used to observe the maze solving as it proceeds to completion. There reporters are enhanced cyan fluorescence protein (ECFP) and enhanced yellow fluorescence protein (EYFP) respectively. The designed TS is an extension of a toggle switch that has been previously built and characterized [5]. The TS is composed of mutually inhibitory repressors, LacI and TetR. Each repressor inhibits the other's transcription, resulting in bi-stability. Exogenous anhydrotetracycline (aTc) or isopropyl-ß-D-thiogalactopyranoside (IPTG) can be used to initialize the TS to a predefined state. Similarly, state changes can be achieved by expression of either repressor via another promoter, such that the current state's promoter
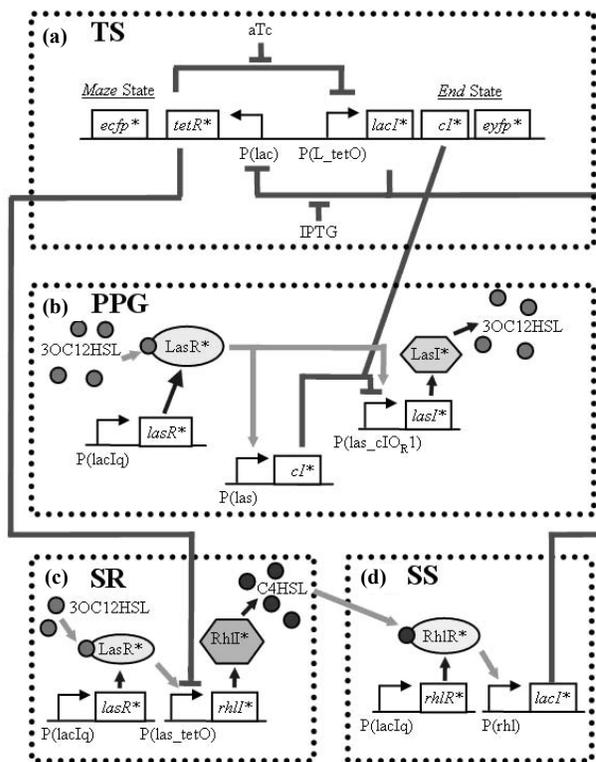
Fig. 3. The interaction between genetic components found in the *End/Maze* cell. The * represents a 12-aa *ssrA* tag (LVA tail) which destabilizes the protein. (a) The TS utilizes mutual repression to achieve bi-stability, which can be either the *End* or *Maze* state. (b) The PPG responds to 3OC12HSL, by transient synthesis of 3OC12HSL, only when the cell is in the *Maze* state. When the cell is in the *End* state, CI represses the PPG. (c) The SR synthesizes C4HSL in response to 3OC12HSL only when the cell is in the *End* state. When the cell is in the *Maze* state, the presence of TetR represses the synthesis of C4HSL. (d) The SS responds to C4HSL by activating the transcription of LacI which forces the TS into the *End* state.

is sufficiently repressed. In addition, CI is transcribed in the *End* state in order to inactivate the PPG.

4) C4HSL Sensor (SS): The C4HSL signal sensor, shown in Fig. 3d, expresses LacI in the presence of C4HSL. LacI subsequently forces the TS into the *End* state.

5) 3OC12HSL to C4HSL Router (SR): The signal router is responsible for the synthesis of C4HSL in response to 3OC12HSL thereby triggering neighboring cells to enter the *End* state (Fig. 3c). The SR is only active in the *End* state, because TetR inhibits expression of RhlI in *Maze* cells.

6) Signal Degrader: The signal degrader (SD) is the sole synthetic circuit in the *Wall* cells. The circuit consists of aiiA and the fluorescence reporter dsRed under control of the P(lacIq) constitutive promoter. The aiiA gene expresses AHL-lactonase, an enzyme that degrades AHLs by hydrolysis [12]. Fig. 3 does not show this genetic circuit because it is not present in the *Maze/End* cell.

III. SIMULATION AND ANALYSIS

A. Simulation

The *End/Maze* cells were modeled using a custom application based on MATLAB 7.0 and Simulink 6.0 (The MathWorks). The components described in the previous sections were designed and analyzed as independent modules with mRNA input/output. The modular modeling technique simplified the simulation procedure by allowing each component to be tested individually before being combined into the complete system.

The *End/Maze* cell was modeled using Hill functions [13]. Equations (1) through (8) show the Hill functions used to describe the interactions in the PPG, the central component of the maze solving system. The equations relate the amount of AHL produced by the PPG in response to AHL diffused in. A similar set of equations was established for each of the components described in the previous section, resulting in a total of 25 equations.

The PPG model includes LasR (R), LasR-AHL complex (RH), CI (C), LasI (I) and AHL (H), with subscripts *m* and *p* representing mRNA and protein respectively. Kinetic constants include translation rates (α_R, α_C, α_I), transcription rates (ψ_I, ψ_C, ψ_R), AHL synthesis rate (ν_H), mRNA decay rates (β_R, β_C, β_I), protein decay rates (γ_R, γ_C, γ_I), binding of LasR to AHL (θ_R) and of the LasR-AHL complex to promoters (θ_{RH}), repression (κ_C), binding cooperativity (η_R, η_C, η_{RH}), and regulatory cooperativity (μ_I, μ_C). Let

$$RH = \frac{R_{PD} \cdot H^{\eta_H}}{(\theta_H)^{\eta_H} + H^{\eta_H}} . \tag{1}$$

Then the time evolution of the simulation is described by:

$$dR_m / dt = \psi_R - \beta_R R_m \tag{2}$$

$$dR_p / dt = \alpha_R R_m - \gamma_R R_p \tag{3}$$

$$dH / dt = \nu_H I_p \tag{4}$$

$$dI_p / dt = \alpha_I \left(\frac{\psi_I \cdot RH^{\eta_{RH}}}{RH^{\mu_I} + (\theta_{RH})^{\eta_{RH}}} \right) \frac{1}{\left(1 + C_m / \kappa_C\right)^{\eta_C}} - \gamma_I I_p \tag{5}$$

$$dI_m / dt = -\beta_I \left(\frac{\psi_I \cdot RH^{\eta_{RH}}}{RH^{\mu_I} + (\theta_{RH})^{\eta_{RH}}} \right) \frac{1}{\left(1 + C_m / \kappa_C\right)^{\eta_C}} \tag{6}$$

$$dC_p / dt = \alpha_C \left(\frac{\psi_C \cdot RH^{\eta_{RH}}}{RH^{\mu_C} + (\theta_{RH})^{\eta_{RH}}} \right) - \gamma_C C_p \tag{7}$$

$$dC_m / dt = -\beta_C \left(\frac{\psi_C \cdot RH^{\eta_{RH}}}{RH^{\mu_C} + (\theta_{RH})^{\eta_{RH}}} \right) \tag{8}$$

Initial values of the constants were obtained from literature [14]-[16], and adjusted to improve the maze solving behavior. Such parameter optimization can be experimentally realized in a forward-engineering project since transcription rates, translation rates, and degradation rates can be altered as necessary through genetic mutations [8] [17].

In order to simulate the complete maze solving system, a spatiotemporal simulator was implemented by integrating MATLAB Simulink models with MATLAB m-files, based on a previous simulator [7]. The spatiotemporal simulator utilizes the single *End/Maze* cell Simulink model, and assumes that colony behavior can be approximated by average cell behavior. The spatiotemporal simulator creates separate Simulink model files for each of the *End/Maze* cells, and differentiates them to the appropriate state. Once the differentiation is complete, the simulator establishes the initial conditions for the simulation and runs the Simulink model for some timestep. The amount of 3OC12HSL and C4HSL produced by each individual colony is then updated on the respective grid. A discrete diffusion equation is then applied to each grid to simulate the diffusion of AHL through agar. Degradation of AHL in *Wall* cells and agar is also accounted for. The values on the grids after diffusion and degradation are then fed back into the Simulink cell models. The cell models are then run again for the same timestep, and the process is repeated until the simulation time is reached. As the timestep approaches zero, the accuracy of the simulation is improved. The discrete diffusion model is shown in equation (9), where *H* is signal concentration in a 2-D agar grid, and Ω is the diffusion constant. The model is an instance of the difference equation and assumes an equal amount of diffusion to the left, right, top, and bottom neighboring grid positions.

$$dH_{(i,j)} / dt = \Omega \cdot \left(H_{(i+1,j)} + H_{(i-1,j)} + H_{(i,j+1)} + H_{(i,j-1)} - 4H_{(i,j)} \right) \tag{9}$$

Several mazes of varying topologies were simulated with the 25 equations describing the complete system using the method described above. The simulations were able to implement all aspects of the designed algorithm. Fig. 4 illustrates the results from simulating one such non-trivial maze. Mazes with loops, bends, and several dead ends were also simulated successfully (data not shown).

B. Sensitivity Analysis

Sensitivity analysis of the PPG was performed in order to identify parameters crucial in governing pulse maximum intensity and propagation delay. The PPG's complex temporal behavior and crucial role as the primary communication device made it a suitable target for sensitivity analysis. In addition, analyzing only the PPG allowed us to focus on a manageable number of parameters. The sensitivity analysis was performed using the Random Sampling High Dimensional Model Representation (RSHDMR) [18] [19]. RSHDMR utilizes the input/output relationship of the system to determine the
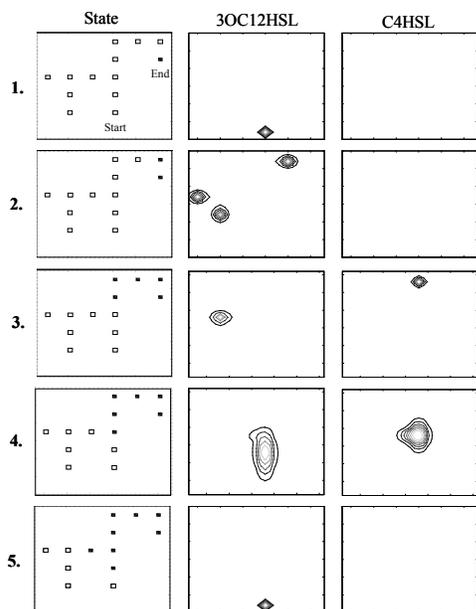
Fig. 4. Maze solving simulation. Each row contains graphs representing the state of the *End/Maze* colonies composing the maze, 3OC12HSL concentration, and C4HSL concentration at specific times. Black squares represent *End* colonies while white squares represent *Maze* colonies. Not shown are the wall cells which surround the maze.

output's dependence on each individual input through function decomposition and high dimensional function fitting.

Ten PPGs were aligned in a row and an exogenous pulse of 3OC12HSL was supplied to the first cell. The input/output map of the system was then developed by running 6,100 iterations of the PPGs with randomly chosen constants. The parameters were varied two orders of magnitude around the nominal values. The pulse max was defined as the maximum AHL concentration in the 8th PPG, and propagation delay as the time between the maximum AHL concentration of the first and last PPG pulse. Each of these traits plays an important role in the functionality of the system; pulse max determines the strength of signal communication between colonies, while propagation delay dictates the time it will take to solve a maze.

The first order sensitivities given by the RSHDMR analysis in Fig. 6a show the top kinetic rate constants in equations (1) through (8) that play an important role in pulse max and propagation delay. Pulse max and propagation delay have the greatest sensitivity to parameters α_I and ψ_I, which control the production of LasI, and therefore the synthesis AHL. Increasing the LasI translation rate (α_I) and transcription rate (ψ_I) results in an increased synthesis of AHL, faster signal propagation, and increased pulse max.

The results from sensitivity analysis provide guidelines for genetic mutations to improve system traits. Each of the parameters discovered to influence pulse max and

	α_I	ψ_I	ψ_C	β_I	γ_I
Pulse Max	1	.83	.55	.48	.67
Propagation delay	1	.98	.2	.19	.51
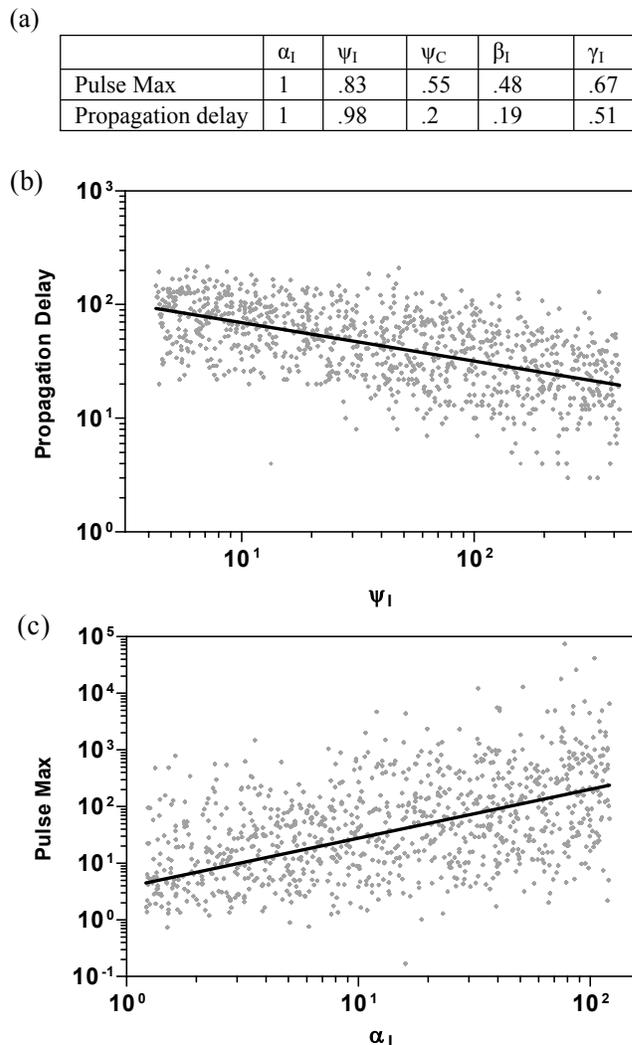
Fig. 6. RSHDMR sensitivity analysis. (a) Table of normalized sensitivity strength of various parameters to pulse gain and propagation delay. (b-c) A single plane of the high dimensional input/output space, where each point represents a single simulation run. Shown is a subset of the runs chosen randomly from the set of 6,100. A linear regression was applied to each log-log plot (P values < .0001). (b) Propagation delay versus ψ_I. (c) Pulse max versus α_I.

propagation delay can then be modulated in the forward-engineering process.

IV. PRELIMINARY EXPERIMENTAL RESULTS

One of the most important and least controlled parameters in the maze-solving system is the AHL diffusion rate. To provide a rough sense of these rates, a *Sender Receiver* experiment was conducted (Fig. 7). Two types of genetically engineered cells were used for the experiment: *Senders* that emit AHL (3OC6HSL) and *Receivers* that respond to the 3OC6HSL by increasing expression of GFP [20] [21]. The *Receiver* cells were plated homogenously on a Petri dish, and a paper disk containing the *Sender* cells was placed in the center. *Receiver* cell fluorescence was then observed at different

(a)

(b)
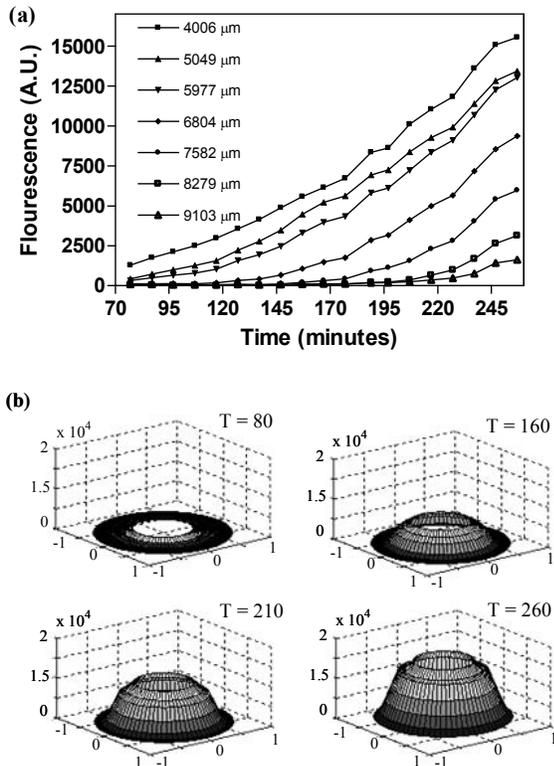
Fig. 7. (a) Graph of fluorescence intensity versus time at various distances from the center of the *Sender* disk. (b) 3-D reconstruction of a 10mm radius of the 35mm Petri dish. The Z-axis represents fluorescence intensity (A.U.) and the x and y axes represent distance away from *Senders* in centimeters. The empty space in the center represents the *Sender* disk. The reconstructions are at 80, 160, 210, and 260 minutes, and were computed by plotting fluorescence intensity (A.U.) versus position and rotating 360 degrees.

times at various distances from the *Sender* cells. The results shown in Fig. 7 can help dictate the dimensions of the maze as well as improve the accuracy of simulations. The diffusion equation (9) utilized in the simulation provides the same characteristic as shown in the experimental results. Work is also being conducted on building and characterizing components, including the PPG, of the maze solving system.

V. SUMMARY

This study shows the first iteration of the engineering process: design, simulation and experimentation. Modeling the biological system as an interconnection of control systems has provided a useful method for simulation and analysis of complex synthetic biological systems. The results from simulation, sensitivity analysis, and experimentation will be applied to the next iteration of the rational design process to achieve an actual maze solving system, which will be the focus of future work.

REFERENCES

[1] E. Ozbudak, M. Thattai, H. Lim, and B. Shraiman, A. Oudenaarden, "Multistability in the lactose utilization network of Escherichia coli," *Nature,* vol. 427, pp. 737-40, 2004.

[2] U. Alon, M. Surette, N. Barkai, and S. Leibler. "Robustness in bacterial chemotaxis," *Nature,* vol. 397, 168-171, 1999.

[3] J. Pomerening, W. Sontag, and J. Ferrell, "Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2," *Nature,* vol. 5, pp. 346-351, 2003.

[4] W. Blake, and F. Isaacs, "Synthetic biology evolves", *TRENDS in Biotechnology,* vol. 22, pp. 321-324, 2004.

[5] T. Gardner, C. Cantor, J. Collins, "Construction of genetic toggle switch in Escherichia coli," *Nature,* vol. 403, 339-342, 2000.

[6] M. Elowitz, and S. Leibler, "A synthetic oscillatory network of transcriptional regulators, *Nature*, vol. 403, pp. 335-338.

[7] S. Basu, R. Mehreja, S. Thiberge, M. Chen, and R. Weiss, "Spatiotemporal control of gene expression with pulse generating networks, *PNAS*, vol. 101, pp. 6355-6360, 2004.

[8] R. Weiss, and S. Basu, "The device physics of cellular logic gates," *The First Workshop on NonSilicon Computing,* pp. 54-61, 2002.

[9] R. Katz, *Contemporary Logic Design.* Redwood City, CA: Addison-Wesley, 1993.

[10] M. Taga, and B. Bassler, "Chemical communication among bacteria," *PNAS*, vol. 100, pp. 14549-14554.

[11] D. Karig, and R. Weiss, "Signal-Amplifying Genetic Circuit Enables *In Vivo* Observation of Weak Promoter Activation in the *Rhl* Quorum Sensing System," *Biotechnology and Bioengineering,* 2005 (In Press).

[12] Y. Lin, J. Xu, J. Hu, L. Wang, S. Ong, J. Leadbetter, and L. Zhang, "Acyl-homoserine lactone acylase from Ralstonia strain XJ12B represents a novel and potent class of quorum-quenching enzymes," *Molecular Microbiology*, vol. 47, pp. 849-860, 2003.

[13] H. Jong, "Modeling and Simulation of Genetic Regulatory Systems," *Journal of Computational Biology*, vol 9, pp. 67-103, 2002.

[14] J. Anderson, C. Sternberg, L. Poulsen, S. Bjorn, M. Givskov, and S. Molin, "New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria," *Applied and Environmental Microbiology*, vol. 64, pp. 2240-2246, 1998.

[15] T. Lederer, M. Takahashi, and W. Hillen, "Thermodynamic Analysis of Tetracycline-Mediated Induction of Tet Repressor by a Quantitative Methhylation Protection Assay," *Analytical Biochemistry*, vol. 232, pp. 190-196, 1995.

[16] X. Zhang, T. Reeder, and R. Schleif, "Transcription Activation Parameters at ara p_{BAD}," *Journal of Molecular* Biology, vol. 258, pp. 14-24, 1996.

[17] Y. Yokobayashi, R. Weiss, and F. Arnold. "Directed evolution of a genetic circuit," *PNAS*, vol. 99, pp. 16587-16591.

[18] G. Li, C. Rosenthal, and H. Rabitz, "High Dimensional Model Representations," *Journal of Physical Chemistry A*, vol. 105, pp. 7766-7777, 2001.

[19] X. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz. "Optimizing genetic circuits by global sensitivity analysis," *Biophysical Journal*, vol. 87, pp. 2195-2202, 2004.

[20] B. Bassler, "Small Talk: Cell-to-Cell Communication in Bacteria," *Cell*, vol. 109, pp. 421-424, 2002.

[21] R. Weiss, and T. Knight, "Engineered Communication for Microbial Robotics," *DNA6*, vol. 2054, pp. 1-16, 2001.