Guest Editorial—Special Issue on Synthetic Biology

Synthetic Biology is an exciting new field that promises to redefine the circuit design, analysis, simulation, and implementation of bio-molecular systems at the DNA, RNA, protein, and small-molecule levels. It has a wide-ranging set of applications in biotechnology, medicine, agriculture, metabolic engineering, energy production, materials synthesis, and bio-inspired computation. Nevertheless fundamental issues with respect to its scaling remain that have prevented its progress over more than a decade. For example: excessive resource consumption that imposes molecular and metabolic toxicity in cells; cross talk; unpredicted or unwanted loading and interactions amongst modules; insufficient attention to signal and cellular stochastics, which are critical in biology; inaccurate models and parameters; a lack of good conceptual methods for design that are inherently robust to error; a lack of attention to analog and probabilistic computation; and the lack of fast simulation tools and design software.

This special issue in Synthetic Biology brings together a set of 10 papers that both review and present original work aimed at addressing these challenges. These papers incorporate advances and insights from analog circuit design, control theory, RNA-focused circuits; from the control and simulation of stochastics (noise) in living cells and their impact on signal detection; from the fast simulation of stochastic and deterministic bio-molecular differential equations via analog and mixed-signal subthreshold circuits; and from mathematical design techniques for synthetic biology.

This journal's traditional focus on bio-inspired and biomedical circuits also makes it ideally suited for reporting advances in the engineering of bio-molecular circuits in living cells themselves, i.e., synthetic biology. Since several papers incorporate both an original and review perspective, **this special issue will provide a good introduction to those unfamiliar with the field as well as provide a rigorous engineering perspective for those actively working in it.** In the following paragraphs, a brief summary of the 10 papers in the order that they appear provides a birds-eye view of the whole special issue.

The papers in this issue are ordered into four main themes: Reviews and big-picture papers of broad applicability to all in the field are presented first, i.e., papers that describe approaches to synthetic biology based on analog circuits, control theory, or RNA-focused circuits respectively. Papers with theoretical and experimental techniques for measuring, analyzing, controlling, or simulating noise in cells and its effect on signal detection are presented next. Papers that report on silicon chips for rapid bio-molecular stochastic or deterministic simulation of cells follow. Finally, papers that focus on mathematical design techniques for synthetic biology occur at the end of the issue.

Digital Object Identifier 10.1109/TBCAS.2015.2472315

I. Analog Circuit Design, Control Theory, and RNA-Focused Circuits

Analog circuit design is the nearly 100-year-old art of crafting and analyzing nonlinear, stochastic, coupled differential equations to perform a desired task, often to given speed, precision, input sensitivity, power, load, or part-count constraints and in the presence of noise or device mismatch. The art uses carefully architected pictorial circuit motifs, symbols, and schematics that allow hierarchical design, analysis, and simulation to be as quantitatively exact or approximate as needed via a whole suite of electronic design techniques and tools. The circuit schematics naturally preserve intuition and feedback-loop relations via pictures, which are essential for analyzing and designing complex systems.

In "Synthetic Biology: A Unifying View and Review Using Analog Circuits" by Teo, *et al.*, the authors map differential equations to pictorial schematics that can represent them quantitatively and thus show how **analog circuit schematics can represent bio-molecular circuits rigorously**. The use of a color-coding scheme reveals that DNA (blue), RNA (red), protein (green), and small-molecule (purple) circuits share many similarities, and **that a canonical analog circuit schematic can represent circuits at all of these levels**. **The authors review over 17 fundamental circuits that have been described in synthetic biology over nearly two decades at all of these levels by constructing analog circuit schematics for them**. These schematics then provide fresh and original circuits-insight into the operation of these bio-molecular systems. Many 'use-it-and-lose-it' feedback loops that are ubiquitous in bio-molecular systems, and which cause 'loading effects' within them, become easy to identify. Many biological circuits are seen as instantiations of similar electronic-circuit topologies such that lessons learned in electronic circuits can be ported to biological circuits and new predictions can be made. **Noise, cellular-machinery and metabolic resource consumption, dynamics, nonlinearity, parameter sensitivity, and loading are automatically and naturally represented, often with easily derivable closed-form expressions**.

Control Theory is a mature mathematical theory for complex-systems design, especially if these systems are linear. As in analog circuit design, which is in many ways, the physical-and-pictorial cousin of mathematical-and-symbolic control-theoretic design, sophisticated feedback systems can achieve desired performance objectives while being robust to certain errors. Such design is often performed with state-space matrix techniques or with transform-based techniques. A well-known control topology uses the 'Proportional-Integral-Derivative' controller or 'PID' controller, which responds in a proportional, integral, or derivative fashion w.r.t. to the feedback error in the topology. PID controllers are also known

as lead or lag controllers in electronic circuit design depending on whether the PD or the PI portion of the control is dominant respectively.

In "Designing Genetic Feedback Controllers" by Harris *et al.*, the authors show how to make linearized or small-signal dynamical models of nonlinear genetic input-output functions. They then use these models to compute Laplace-transform-based transfer functions. Such transfer functions are then combined in a negative-feedback loop to design a genetic PID controller, specifically a lag controller. The lag controller is implemented by having a relatively low-pass two-molecule genetic cascade pathway and a one-molecule genetic pathway effectively add by creation of a common output molecule. The common output molecule also serves as a protein-digesting molecule (a 'protease') that can specifically digest the input molecule of both pathways and thus remove it. The removal of the input molecule by the protease leads to a negative-feedback loop wherein the output molecule reduces the effect of the constitutively produced input molecule. **Thus, the overall system implements a lag-controlled bio-molecular system around a linearized operating point**. This work is an excellent example of a **growing trend in synthetic biology today to replace digital logic-gate-based designs with more rigorous control-theory-based designs**.

Circuit design based on transcriptional regulation of DNA promoters is dominant in synthetic biology. Increasingly, **translational regulation of RNA is being recognized as critical** because the synthesis of proteins affects metabolic loading, the use of common ribosomal resources, and osmotic toxicity in the cell, all of which have limited the scaling of synthetic biology. **In "Multilevel Regulation and Translational Switches in Synthetic Biology" by Kopniczky *et al.***, devices and circuits based on translational regulation are reviewed including those that use 'MotBPs', i.e., RNA Motif Binding Proteins, and those that use signal-inducible riboswitches.

The latter review points out three major benefits to translational control and emphasizes its importance in mammalian synthetic biology: i) Since two input signals can simultaneously control the production of one output protein, with one input regulating output translation, and the other input regulating output transcription, multi-level circuits that are part-count efficient can be built. The use of RNA to perform multiple-input logic similar to that done at DNA promoters, can further improve efficiency. Circuits that use such strategies have been used to construct efficient XOR and AND operations needed in a half adder using very few parts and form part of the review. ii) The derepression of mRNA translation is typically much faster than the derepression of DNA transcription such that translational control can be significantly faster than transcriptional control for the delay associated with the onset of molecular production. iii) Molecular inputs that alter the secondary structure of mRNA, the exposure of ribosome binding sites, the configuration and location of terminator or anti-terminator loops for RNA polymerase, the recruitment of translation initiation factors, the steric blocking of ribosomes, or the degradation of mRNA all serve to create translation-control-based sensors or computing elements.

II. Noise and Signal Detection

The 'excess protein noise' seen in cells is often quantified by the 'Fano factor'. It is largely due to mRNA noise that is amplified by translation that adds to the noise generated by the process of translation itself. **In "Controlling *E. coli* Gene Expression Noise" by Kim *et al.***, the authors use different spacers between the ribosome binding site and the start codon in *E. coli* to vary translation gain by almost 60x, and elegantly control the mean, noise, and Fano factor of an output fluorescent protein. At low output protein copy numbers and low translational gain, they observe the noise due to translation itself. These measurements of noise in cells also correct for auto-fluorescence effects at low protein copy numbers via deconvolution. After such correction, **the authors show impressive agreement between theoretical predictions of gamma distributions of cell numbers vs. fluorescence and experimental measurements of these same distributions**. Their prior work on identifying the right parameters to sensitively control desired outputs enables principled methods for varying the noise vs. the mean of the copy number of a given protein. Therefore, the authors also show good agreement between theoretical predictions of the protein-copy-number variance and mean vs. experimental measurements of these same quantities. **The work in the latter paper should be of great use to several synthetic biologists interested in understanding and controlling intrinsic noise in cells, particularly in bacterial cells**.

In "Efficient Sampling of Bacterial Signal Transduction for Detection of Pulse-Amplitude Modulated Molecular Signals" by Bicen *et al.*, the authors investigate the detection of Pulse Amplitude Modulated (PAM) detection of a molecular input inducer by bacteria. The molecular input activates an output fluorescent response that is well characterized by classic Michaelis-Menten-like protein production and a linear decay. An experimental microfluidic setup enables flow-rate and amplitude control of the input as well as optical reporting from the bacteria, which reside in an indented chamber within the microfluidic setup. The pulsatile input leads to a pulsatile output with time-dependent signal and noise statistics, both of which are measured. Based on these measurements, the authors show that **the ramp-up slope of the bacterial response offers a good measure of the signal amplitude with a low probability of error**. The latter measure significantly outperforms other measures such as the peak time of response, the total response duration, or the ramp-down slope with a probability of error that is at least an order-of-magnitude lower than that of all other measures. This work will be useful to all synthetic biologists who are interested in engineering bacteria to detect signals in their environment, an application of great value in many biotechnology and security applications.

The exact stochastic dynamic simulation of even low-order biochemical differential equations via the 'Chemical Master Equation' developed by Gillespie and others in incredibly computationally intensive: For example, in **"Conditional Moment Closure Schemes for Studying Stochastic Dynamics of Genetic Circuits" by Soltani *et al.***, for just a simple two-gene circuit with an activator and repressor in a negative-feedback loop, the authors point out that 20,000

Monte-Carlo simulations in MATLAB are needed to obtain quantitative information about the mean, variance, and other moments of the joint probability distributions of the four protein and gene variables. On a computer with 8 GB RAM and four 3.4 GHz cores, such simulations take 20 minutes. It is likely that the stochastic simulation of an entire *E. coli* cell on a supercomputer with 30,000 state variables for just one cell cycle ($\sim$20 min) would take several years on a supercomputer. Therefore, there is great motivation to either simplify these equations or to build custom silicon chips capable of vastly speeding up such stochastic simulations. In the next section, we shall discuss custom chips that are designed for fast simulations. In this section, we describe how the latter paper utilizes mathematically rigorous techniques for simplifying the chemical master equation.

In the latter paper, the exact stochastic differential equations are mapped to differential equations that describe the dynamics of relatively-low-order moments of the underlying probability distribution, e.g., the mean, variance, skewness, correlations, third-order moments etc. While such techniques do not provide the full probability distributions, they are nevertheless useful for design in synthetic biology: they can provide quantitative information on how the noise in cellular circuits depends on various parameters, quickly, e.g., in less than a second versus 20 min. for a particular two-gene circuit. A problem with such approaches is that higher-order moments, which affect the time derivatives of the lower-order moments, need to be approximated as a function of the lower-order moments, or the overall simplified differential equations on the low-order moments will have gross errors, and not converge. **The authors describe a 'Conditional Derivative Matching' (CDM) method that yields simple moment equations that are accurate and that converge where other methods fail.** For example, they show that neither Gaussian approximations to the probability distribution that ignore all moments beyond the second moment, nor unconditional derivative matching, which are used to generate equations for the higher-order moments as a function of the lower-order moments are adequate. They are inadequate for even the two-gene negative-feedback circuit particularly at low copy numbers. **The key insight is to condition protein levels on active/inactive states of randomly toggling genes and then express higher-order moments as functions of lower-order conditional moments**. The equations and techniques in the latter paper should be useful to several synthetic biologists designing low-copy-number circuits, which are essential if we want to prevent osmotic and metabolic toxicity in cells and yet have complex circuits with many parts.

III. Cytomorphic Chips for Rapid Stochastic and Deterministic Simulations

There are astounding similarities between the equations that describe noisy electronic flow in sub-threshold transistors and the equations that describe noisy molecular flow in chemical reactions, both of which obey the Boltzmann laws of exponential thermodynamics. Based on these similarities, it is possible to take a principled ***cytomorphic*** approach to design log-domain circuits in living cells and to simulate them via subthreshold analog circuits. For example, prior work by the authors has engineered logarithmic analog computation in *E. coli* with less than three transcription factors, almost two orders of magnitude more efficient than prior digital approaches to create a 'bio-molecular slide rule'. **In addition, highly computationally intensive noisy DNA-protein and protein-protein networks can be rapidly simulated in mixed-signal supercomputing chips that naturally capture their noisiness, nonlinearity, dynamics, and non-modular interactions**. Even physical temperature in biological circuits maps to the physical temperature in cytomorphic circuits.

In "A Cytomorphic Chip for Quantitative Modeling of Fundamental Bio-Molecular Circuits" by Woo *et al.*, the authors describe a 0.35 μm **digitally programmable analog BiCMOS silicon chip** capable of simulating all fundamental biochemical basis functions, and of being scaled and composed into larger networks via the use of digital molecular address and data packets. The fundamental basis functions include general biochemical binding with loading in cascade and fanout topologies; programmable Hill-coefficient and cooperative binding; inducer, transcription factor, and DNA binding; probabilistic gene transcription with analogic log-linear and saturating operation; gain and degradation of mRNA and protein dynamics; and, stochastic circuits for faithfully representing highly noisy biological circuits via tunably noisy analog transistor circuits. Published MATLAB models that describe **biological data are in very good agreement with chip simulations of these same models and also with the biological data. Computationally intensive Gillespie simulations are also faithfully and rapidly reproduced by the chip and reliably tunable over the range of signal-to-noise ratios observed in biological cells**. A relatively wide dynamic range of operation (100 dB+) and extensive digital programmability promise that **cytomorphic approaches, when scaled, will enable rapid and large-scale simulations of stochastic and deterministic bio-molecular networks in the future**.

In "A 1.26 μW Cytomimetic IC Emulating Complex Nonlinear Mammalian Cell Cycle Dynamics: Synthesis, Simulation, and Proof-of-Concept Measured Results" by Houssein *et al.*, continuing on the same theme, **log-domain subthreshold circuits are used to simulate a network of cyclin-dependent kinases driving the mammalian cell-cycle network**. The cell cycle is extremely important in regulating cell division and differentiation. Malfunctions in the cell-cycle pathway, particularly in checkpoint proteins that regulate transitions between its states are often implicated in cancer. The cell cycle is composed of four primary states, G1 (growth), S (DNA replication), G2 (more growth), and Mitosis (Cell division), along with a G0 quiescent phase, that is entered into when cells are 'senescent'. **The authors use a published 5-state-variable set of coupled deterministic differential equations** that are a simplification of a more elaborate published 39-state-variable model as the basis for their work. They use a 'Nonlinear Bernoulli Cell Formalism', which is well known in log-domain circuit design, to map a 5-state-variable biological set of differential equations into equivalent log-domain subthreshold circuit equivalents on a proof-of-concept 0.35 μm chip. They verify successful operation via measured chip results and via simulations. **The authors show that their nonlinear cell-cycle circuit can**

exhibit multiple oscillatory behaviors including simple oscillations, complex oscillations, quasi periodicity, and even chaos. One note-worthy feature of their chip, as in many sub-threshold circuits, **is the low 1.26 μW power consumption**. The low power consumption suggests that large-scale portable cytomorphic chips may be used in implantable cell-sensing and cell-control medical-device applications of the future.

IV. Mathematical Design Techniques for Synthetic Biology

Traditional metabolic engineering of cellular pathways is founded on the principles that: i) Taking the stoichiometry of the metabolites that participate in each biochemical reaction in cells into account, the biochemical reaction fluxes must be constrained such that each and every metabolite balances at steady state, i.e., the reaction fluxes that create or consume each metabolite balance; ii) each reaction flux is within certain physiological and plausible bounds; and, iii) a certain weighted function of the fluxes is optimized by the cell. The weighted function of the fluxes might be an objective function of biological relevance to the cell, e.g., **a configuration of reaction fluxes that maximizes the cell's biomass from given nutrients, a given genetic makeup, and given enzyme levels**. In addition, a synthetic objective function of the fluxes may attempt to **maximize production of a metabolite of commercial interest**, e.g., the production of 1,4 Butadienol (BDO) in *E. coli*, useful as an industrial solvent, and in the manufacture of elastic fibers and plastics. **The synthetic and natural objective functions exhibit tradeoffs with respect to each other and effectively define a multi-objective optimization problem. This multi-objective optimization problem can be addressed in a powerful way using a mathematical concept known as Pareto optimality**. Pareto optimality was first developed in economics, and is now used in engineering as well. The Pareto-optimal fronts define a mathematically rigorous way of quantifying and achieving tradeoffs between objective functions. Manipulations such as engineered gene knockouts with associated costs and that up-regulate or down-regulate enzyme activity in the reactions affect different objective functions in different ways. **In "Pareto Optimal Design for Synthetic Biology" by Patane *et al.*, the authors use strict and relaxed variants of Pareto optimality to show that they can significantly outperform previous optimization approaches for five synthetic metabolites including BDO and certain organic-acid biosynthesis important in biofuels**. This work could be quite important in bringing new optimization techniques into the field of metabolic engineering, an important sub field within the larger field of synthetic biology.

In "Designing Conservation Relations in Layered Synthetic Biomolecular Networks" by Prescott *et al.*, the authors suggest a design paradigm for synthetic circuits in cells wherein fast reactions in a 'fast-reaction layer' can architect designed polynomial conservation laws that effectively determine relations between slow state variables in a 'slow-reaction layer'. The fast-reaction layer may arise from a Protein Interaction Network (PIN), while the slow reaction layer may arise from a Genetic Regulatory Network (GRN). One of the key ideas suggested in the latter paper is that, if the stoichiometry of fast reactions is engineered appropriately, forward-reaction and reverse-reaction equilibrium in the fast reactions will automatically lead to polynomial conservation equations for slow variables, which are based on mass-action equations. These laws will hold for all times beyond a fast transient. The authors show that a given conservation relation can be architected in several ways in the fast layer, e.g., in one reaction via dimerization to yield a square-law conservation relation; or, via a sequence of two reactions to also yield a square-law conservation relation. **Thus, state variables in the slow layer may appear to be interacting but it is really because the fast layer is making it appear that they do**. Techniques for time-scale separation are extremely important in stabilizing feedback loops in electronic circuits. Time-scale-separation concepts such as those in the latter paper, may impact the design of cell-inspired electronic circuits in the future as well.

We hope that the readers of this special issue will enjoy a fascinating collection of papers.

Rahul Sarpeshkar, *Thomas E. Kurtz Professor*
Professor of Engineering, Physics, Microbiology & Immunology, and Physiology & Neurobiology
Dartmouth College
Hanover, NH USA

Rahul Sarpeshkar received B.S. degrees in electrical engineering and physics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, and the Ph.D. degree from the California Institute of Technology, Pasadena, CA, USA.

Currently, he is the Thomas E. Kurtz Professor at Dartmouth College, Hanover, NH, USA, where he is also a Professor in the departments of Engineering, Physics, Microbiology and Immunology, and Physiology and Neurobiology. His research creates novel wet DNA-protein circuits in living cells and also advanced dry nano-electronic circuits on silicon chips. His longstanding work on analog and biological computation and his most recent work have helped pioneer the field of analog synthetic biology. His work on a glucose fuel cell for medical implants was featured by Scientific American among 2012's 10 World Changing Ideas. He holds over 35 awarded patents and has authored more than 125 publications, including one that was featured on the cover of Nature. His book, *Ultra Low Power Bioelectronics: Fundamentals, Biomedical Applications, and Bio-Inspired Systems*, revealed the deep connections between analog transistor circuits and biochemical circuits and founded the field of cytomorphic systems. His group holds several first or best records in analog, bio-inspired, synthetic biology, medical device, ultra low power, and energy harvesting systems. His work has applications in implantable medical devices for the deaf, blind, and paralyzed and in biotechnology and medical applications that benefit from cellular engineering. Before he joined Dartmouth's faculty, he was a tenured professor at MIT, leading the analog circuits and biological systems group at the Research Lab of Electronics. Before joining MIT, he was a member of the technical staff of Bell Labs' division of biological computation in their physics department.

Dr. Sarpeshkar has received several awards including the NSF Career Award, the ONR Young Investigator Award, and the Packard Fellows Award.