Energy Aware Technology Mapping of Genetic Logic Circuits

Erik Kubaczka,[†] Maximilian Gehri,[†] Jérémie J. M. Marlhens,[†,‡] Tobias Schwarz,[†] Maik Molderings,[†,‡] Nicolai Engelmann,[†] Christian Hochberger,[†,¶] and Heinz Koeppl[*,†,¶]

†*Department of Electrical Engineering and Information Technology, TU Darmstadt, Darmstadt, 64283, Germany*

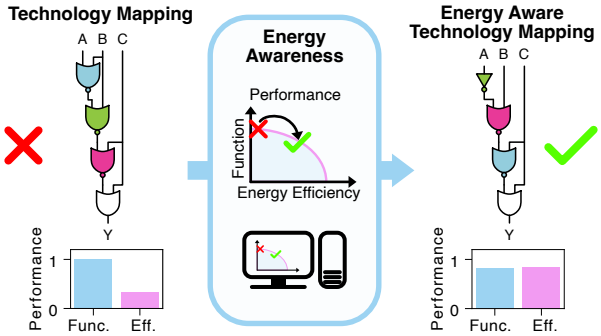
‡*Graduate School Life Science Engineering, TU Darmstadt, Darmstadt, 64283, Germany*

¶*Centre for Synthetic Biology, TU Darmstadt, Darmstadt, 64283, Germany*

E-mail: heinz.koeppl@tu-darmstadt.de

IWBDA 2023 Special Issue

Abstract



Energy and its dissipation are fundamental to all living systems, including cells. Insufficient abundance of energy carriers -as caused by the additional burden of artificial genetic circuits- shifts a cell's priority to survival, also harming the functionality of

the genetic circuit. Moreover, recent works have shown the importance of energy expenditure in information transmission. Despite living organisms being non-equilibrium systems, non-equilibrium models capable of accounting for energy dissipation and non-equilibrium response curves are not yet employed in genetic design automation (GDA) software. To this end, we introduce Energy Aware Technology Mapping, the automated design of genetic logic circuits with respect to energy efficiency and functionality. The basis for this is an energy aware non-equilibrium steady state (NESS) model of gene expression, capturing characteristics like energy dissipation -which we link to the entropy production rate- and transcriptional bursting, relevant to eukaryotes as well as prokaryotes. Our evaluation shows that a genetic logic circuit's functional performance and energy efficiency are disjoint optimization goals. For our benchmark, energy efficiency improves by 37.2% on average when comparing to functionally optimized variants. We discover a linear increase in energy expenditure and overall protein expression with the circuit size, where Energy Aware Technology Mapping allows for designing genetic logic circuits with the energy efficiency of circuits that are one to two gates smaller. Structural variants improve this further, while results show the Pareto dominance among structures of a single Boolean function. By incorporating energy demand into the design, Energy Aware Technology Mapping enables energy efficiency by design. This extends current GDA tools and complements approaches coping with burden *in vivo*.

Keywords

Genetic Design Automation; Energy; Non-Equilibrium; Thermodynamics; Synthetic Biology; Gene-Expression; Technology Mapping; Metabolic Burden; Computer Aided Design; Entropy Production Rate

1 Introduction

Life is non-equilibrium (*1*), and so energy and its dissipation is essential for life and the function of biological organisms and systems (*2*, *3*). While the absence of energy is incompatible with living (*1*, *2*), a decrease in it's availability has detrimental effects on a cell's metabolism and in consequence on it's growth, fitness, and gene expression (*4–9*). Such a decrease can be sourced in insufficient nutrition (*4*) of cells but also in the burden imposed by the insertion of engineered genetic circuits (*8*) and the expression of heterologeous proteins (*5*). Metabolic burden, observed in both *S. cerevisiae* and *E. coli*, refers to the diversion of resources from the host to synthetic constructs, impacting energy, nutrients, ribosomes, and RNA polymerase, as well as reducing cellular fitness (*5*, *6*). This allocation away from maintenance and growth compromises the host organism's physiological functions. As the functionality of synthetic constructs such as engineered genetic circuits depends on sufficient dynamics of proteins and other molecules (*10*) a reliable and well functioning host organisms are essential (*7*). Consequently, a trade-off between function and energy efficiency emerges, affecting reaction levels even at the promoter scale (*11–13*).

The described importance of energy did not hinder the wide application of equilibrium gene expression models (*14–17*). Raised in the vicinity of bacterial transcription, these models assume that regulatory mechanisms, such as transcription factor binding to DNA, operate at thermodynamic equilibrium (*14–16*, *18*). While this seems reasonable for prokaryotes, non-equilibrium processes and the inherent energy dissipation are essential for gene regulation in eukaryotes (*1*, *3*, *18–20*). In particular the sharpness and sensitivity of the eukaryotic gene response surpasses the so called Hopfield barriers (*3*, *11*, *20*), which place an upper limit in the equilibrium case and where first encountered in the context of kinetic proofreading by J. Hopfield (*21*). Besides this, transcriptional bursting is characteristic for eukaryotic gene expression (*11*, *22–25*) but is also present in prokaryotes (*26*). These bursts are characterized by durations of transcriptional activity significantly longer than the binding of single transcription factors, which lasts a few seconds (*20*, *23*, *25*). Mechanisms for this may in-

clude multistep activation, where transcription factor abundance modulates transcriptional activity, or cooperative exchange, where the burst period is determined by transcription factors rapidly swapping positions due to cooperative binding (22, 23, 25).

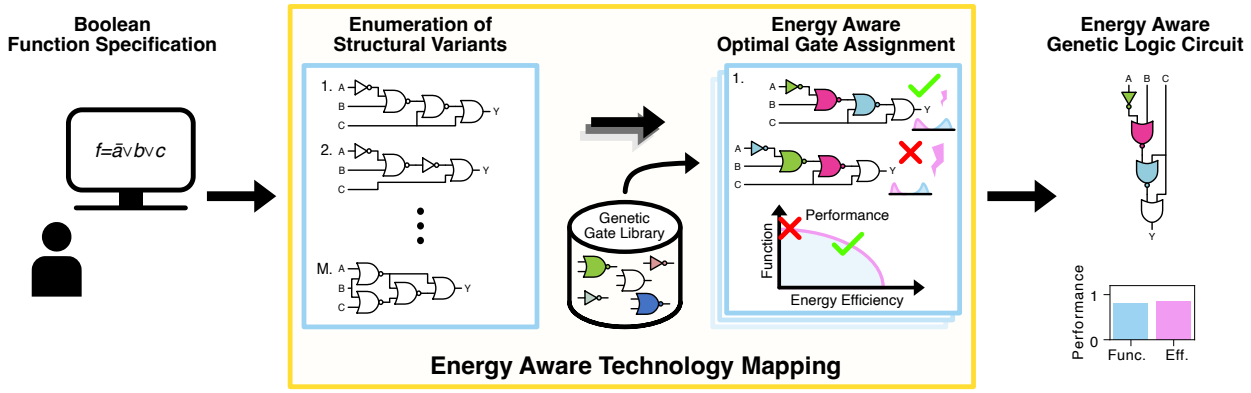


Figure 1: **Energy Aware Technology Mapping.** We here present the technology mapping pipeline at the example of the proposed Energy Aware Technology Mapping. With the specification of the Boolean function to realize as input, the technology mapping first enumerates structural variants (27). For each circuit structure obtained, we perform *in silico* an energy aware gate assignment. This takes into account the genetic logic circuit's performance with respect to both, energy efficiency and functional performance. After successful completion of the process, the user receives the automatically designed genetic logic circuit.

Despite these challenges, the targeted engineering of biological systems advances rapidly with standards and tools aiding in or automating their design being created (10, 28–33), while the awareness with respect to resources such as energy increases (8, 9, 34–36). One branch of tools is Genetic Design Automation (GDA) software (10, 33, 37, 38). These tools solve the task of creating genetic logic circuits realizing Boolean functions from modules characterized in a gate library (10, 37, 38), the so called technology mapping. GDA tools optimize the characteristics of genetic logic circuits by evaluating them with the use of models on the basis of *in silico* experiments. The primary objective is the functionality, which expresses the circuits capabilities of implementing the Boolean function in terms of a distance related to the minimum fold-change between two distinct Boolean states (10, 27). To be meaningful, the models employed have to capture the characteristics inherent to the cells and constructs under consideration, such as the input-output characteristics of gene

expression cascades, non-equilibrium attributes like energy dissipation and the inclusion of transcriptional bursting, being characteristic for eukaryotes (*11*, *22*, *23*, *25*, *39*) but also encountered in prokaryotes (*26*).

Technology mapping is originally one step during the hardware synthesis of digital systems. It takes a functional specification and maps it to elements of the target technology. In the case of GDA, the target technology usually provides genetic logic gates which are collected in a library, including multiple different implementations of the same gate type. In comparison to hardware synthesis, every specific gate realization can only be used once in such a circuit, while in most cases not all gate types are available. Thus, a suitable topology of gates has to be found which exhibits the specified behavior. Multiple different structures can be found that fulfill this requirement. The assignment of gates from the library to the elements of the topology completes the creation of the genetic logic circuit.

Otero-Muras et al. (*40*) present a tool to automatize the engineering of synthetic metabolic pathways on the basis of Pareto optimal designs. These metabolic pathways are defined by the user, with genetic logic circuits being a possible branch. iBioSim (*33*, *41*) is a tool for the automatized construction of genetic circuits, their simulation and model representation. The absence of an automatized flow from Boolean specifications to a DNA sequence highlights, that this tool does not primarily target the automated design of genetic logic circuits. In contrast to iBioSim, Genetech (*42*) starts from a two level Boolean expression and transforms this expression to match the NOT and NOR gate structure of the Cello library. Although enumerating all possible gate assignments, their functionality is not assessed. Cello (*10*, *37*, *38*) probably has the best user interface for engineering genetic logic circuits. The user can specify the desired boolean function using the hardware design language Verilog. The given function is then transformed to a functional equivalent circuit which only uses elements of the provided libraries, which exist for S. cerevisiae and E. coli. The gate assignment is optimized with respect to the smallest fold change between circuit output values corresponding to the distinct Boolean states ON and OFF, the cello-

score. ARCTIC (*17*, *27*) extends the focus on the robustness of the resulting genetic logic circuits. By presenting particle based simulation, the E-score, and structural variants, this tool accounts for the stochastic nature of genetic gates and adds topologies to the design space of GDA. In the 2023 version of this tool, context awareness was introduced into the tool. This can refer to unwanted mutual influences of gates (which were supposed to have no interaction) or interaction of the proteins that are used as signal carriers with the host.

While the awareness of energy in special and resources in general increases (*8*, *9*, *34–36*), it is not yet part of GDA. The same holds for non-equilibrium models of gene expression, accounting for energy expenditure and sharpness beyond the Hopfield barriers. To this end, we introduce **Energy Aware Technology Mapping**, the design of genetic logic circuits with respect to function and energy. For this, we first present a probabilistic non-equilibrium steady state (NESS) model of gene expression on the basis of a continuous time Markov chain (CTMC). This model features a modular design, which can account for promoter architectures varying in the number of binding sites, activation steps, and cognate transcription factors. The number of bound transcription factors does not determine transcriptional activity. Rather, it modulates the switching rates between different transcriptional activities and allows for transcriptional bursting in the multistep activation as well as the exchange model (*11*, *22–26*). Additionally, the model can capture non-equilibrium characteristics like energy dissipation, which is essential to all living organisms (*2*, *3*). In this context, we relate the entropy production rate of our model to its thermodynamic energy dissipation rate.

By giving rise to the energy dissipation and functional response characteristic realized by the gene, the introduced NESS gene expression model serves as basis on which we establish Energy Aware Technology Mapping, the design of genetic logic circuits with respect to function and energy. This allows us to explore the trade-off between energy efficiency, which we consider as reciprocal of the energy demand, and function for genetic logic circuits. To this end, we first consider the Boolean circuits presented in (*10*) and continue with the evaluation of the impact of structural variants and the associated Pareto frontiers. The

Pareto frontiers give rise to the structure's performance independently of the optimization objective considered. By presenting means for multi-objective optimization, we allow for trading-off the objectives function and energy efficiency in a joint optimization.

The results and methods we present here are implemented in the technology mapping framework ARCTIC. You find ARCTIC at https://www.rs.tu-darmstadt.de/ARCTIC.

2 Results

2.1 An Energy Aware Gene Expression Model

Gene expression is an inherently non-equilibrium process (*1*, *43*), depending on the presence of energy carriers and building blocks like ATP and charged tRNA. While gene expression is subject to regulation at various levels (*3*, *44*, *45*), we focus on transcriptional regulation through the binding of activating or inhibiting transcription factors. In this section, we present the response characteristics of a single protein-coding gene whose promoter has one or more binding sites for cognate transcription factors. To this end, we describe the dynamics of the promoter and of the abundances of its associated mRNA and protein via a stochastic chemical reaction network (CRN). In contrast to combinatorial (equilibrium) promoter models, the use of a kinetic model enables the description at non-equilibrium. In particular, we obtain the mean energy dissipation rate and the NESS mean and variance of protein abundance as a function of the transcription factor concentration. These non-equilibrium response curves are used to characterize the genes realizing the gates for the technology mapping and posses attributes like increased sharpness in comparison to their equilibrium counterparts (*11*). We will first elucidate the kinetic model and then introduce the thermodynamic concepts necessary to relate the kinetic model to its energy dissipation rate.

A CRN is composed of the chemical species $\mathsf{X}_1, \ldots, \mathsf{X}_N$, and reactions $\mathcal{R}_1, \ldots, \mathcal{R}_M$ with

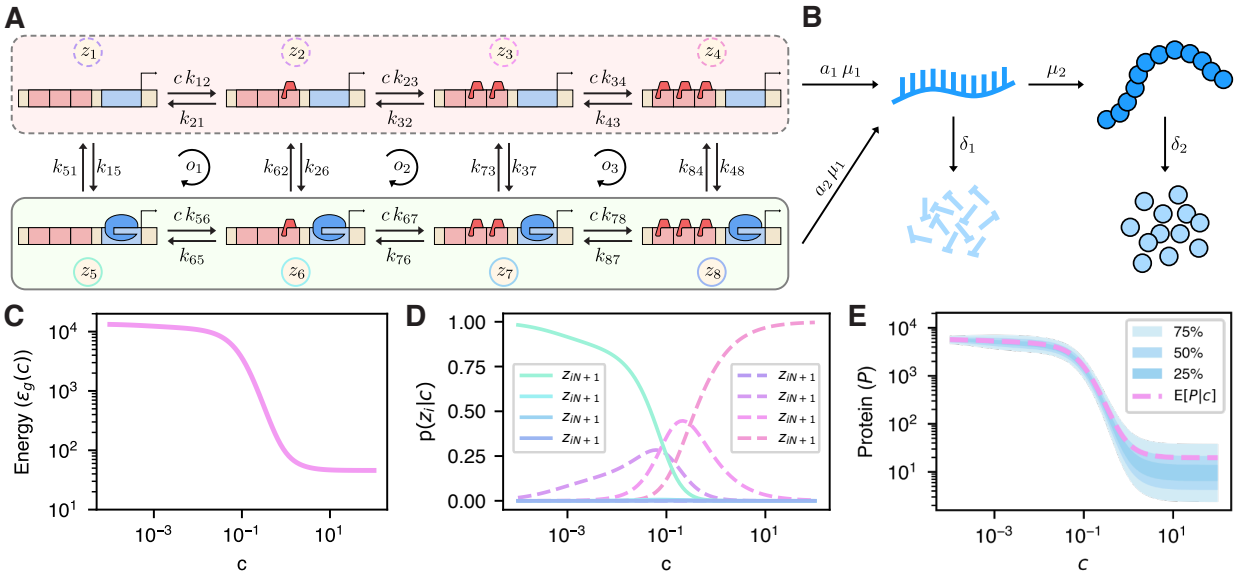


Figure 2: **Energy aware gene expression model. AB:** Schematic description of the proposed model consisting of the promoter model (**A**) and the reactions describing the RNA and protein dynamics (**B**). **A:** The promoter model is instantiated with two levels of transcriptional activity and capabilities of binding up to three transcription factors. The transcription factor concentration enters via the variable c, with the transcriptional active states (z_i with $i = 5, 6, 7, 8$) featuring transcription rate $a_2\,\mu_1$ and the inactive states (z_i with $i = 1, 2, 3, 4$) the basal rate $a_1\,\mu_1$ ($a_2 > a_1$). **B:** Besides transcription, the dynamics include translation (μ_2) and the respective degradation reactions (δ_1 and δ_2). **C-E:** These figures present an exemplary response characteristic of the model in dependence of the transcription factor abundance c. We here showcase inhibitory behaviour of the transcription factor, while the model can also capture activatory behaviour. **D** visualizes the steady state probabilities of each state, which indicates that upon increasing c the probability mass shifts from z_5 to z_2 and z_3 before concentrating in z_4. The corresponding protein distribution is shown in **E** by the mean and three quantile intevals, while **C** presents the expected energy dissipation rate of the overall model. Comparing **C** and **E**, one notices the proportionality between energy dissipation rate and protein abundance.

stoichiometric balanced equations

$$\mathcal{R}_m : \quad \sum_{i=1}^{N} u_i^{(m)} \mathsf{X}_i \quad \underset{\tilde{k}_m^-}{\overset{\tilde{k}_m^+}{\rightleftharpoons}} \quad \sum_{i=1}^{N} w_i^{(m)} \mathsf{X}_n, \tag{1}$$

where the backward reaction microscopically reverses the forward reaction. We assume the reactions to follow the stochastic law of mass action (*46*), which assigns the propensity functions $\lambda_m^+(\boldsymbol{x})$, $\lambda_m^-(\boldsymbol{x})$ to each reaction $\mathcal{R}_m$. These describe the rate at which the reaction occurs when the system is in state $\boldsymbol{x}$, while $\tilde{k}_m^+ > 0$ and $\tilde{k}_m^- \geq 0$ are the corresponding reaction rate constants. A reaction is called *microscopically reversible* if $\tilde{k}_m^- > 0$. The concept of microscopic reversibility is important for the later thermodynamic treatment.

We partition the vector $\boldsymbol{X}$ of random variables into the promoter $\boldsymbol{Z}$ and the RNA and protein abundances R and P. It is assumed that the promoter can switch microscopically reversibly between a finite number of states z_i, where we assign the transcriptional activity a_i to each state. These states are represented as different species, so that $\boldsymbol{Z}$ is a unit vector with a one at position i in each case. RNA synthesis, protein synthesis, and degradation are modelled as single reactions, neglecting their multistep construction (*2*). Although each of the elementary steps is microscopically reversible, we assume that the rate constant of the mechanistically reversed reaction is negligibly small, enforcing microscopic irreversibility. In Figures 2**A**, **B** we represent an exemplary instantiation of the gene expression model, where the promoter (**A**) is parameterized to feature three binding sites and two distinct levels of transcriptional activity. Given the concentration vectors of the transcription factors $\boldsymbol{c}$, the transition rates of the promoter (as in Figure 2**A**) relate to Equation (1) as $\tilde{k}_m^\rho = f_{ij}(\boldsymbol{c}) k_{ij}$ for $\boldsymbol{\nu}_m = \boldsymbol{e}_j - \boldsymbol{e}_i$, where $\rho = +$ for $j > i$ and $\rho = -$ for $j < i$. The functions f_{ij} describe the dependence of the transition rate constant from state z_i to z_j on the transcription factor concentrations. The CRN representation of a promoter model is more versatile in general, as it may describe not only the number of bound transcription factors but also different DNA conformations in a potentially multi-step transcriptional activation (*11*). Such conformations

can be, for example, open and closed loop complexes or different chromatin states (*11*, *19*, *47*). It should be noted, however, that some of these activation steps may be ATP-dependent in one direction but not in the other. In order to maintain microscopic reversibility of all promoter state changes, which is a fundamental assumption for our expression of the energy demand of the promoter, activation and deactivation may thus be modeled via distinct reactions. Also, a change in promoter state must not occur simultaneously with RNA synthesis, which prohibits the use of a multi-step transcription initiation model featuring different RNA polymerase complexes at the promoter.

The evolution of the distribution of the state of the gene expression system $\mathrm{p}(\boldsymbol{X}(t) = \boldsymbol{x})$ is described by the chemical master equation given in SI Section S1.3. Using the Chapman-Kolmogorov backward equation, we derive the mean and variance of the non-equilibrium steady state RNA and protein distributions in SI section S2, while the steady state promoter distribution $\boldsymbol{\pi}(\boldsymbol{c}) = \lim_{t\to\infty} \mathrm{p}(\boldsymbol{Z}(t) = \boldsymbol{e}_i)$ was derived using the methods presented in section 4.2 via the propensity matrix $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{c})$ with $\Lambda_{ij}(\boldsymbol{c}) = f_{ij}(\boldsymbol{c})k_{ij}$. Together with the transcription activity parameter a_i associated with each promoter state z_i, the average promoter activity is

$$\alpha(\boldsymbol{c}) = \sum_{i=1}^{n} a_i\, \pi_i(\boldsymbol{c}). \tag{2}$$

The protein's mean and variance and the RNA's mean are given by

$$\mathsf{E}[P \mid \boldsymbol{c}] = \frac{\mu_2}{\delta_2}\, \mathsf{E}[R \mid \boldsymbol{c}] \qquad\qquad \mathsf{E}[R \mid \boldsymbol{c}] = \frac{\mu_1}{\delta_1}\, \alpha(\boldsymbol{c}) \tag{3}$$

$$\begin{aligned}\mathsf{Var}[P \mid \boldsymbol{c}] = {} & \frac{\mu}{\delta}\, \alpha(\boldsymbol{c})\, \left(1 - \frac{\mu}{\delta}\, \alpha(\boldsymbol{c})\right) + \frac{\mu_2}{\delta_2 + \delta_1}\, \frac{\mu}{\delta}\, \alpha(\boldsymbol{c}) \\ & + \frac{\mu^2}{\delta + \delta_2^2}\, \boldsymbol{a}\, \left(\delta_1^{-1} + \boldsymbol{M}_2^{-1}\right)\, \boldsymbol{M}_1^{-1}\, (\boldsymbol{a} \odot \boldsymbol{\pi}(\boldsymbol{c}))\end{aligned} \tag{4}$$

where $\mu = \mu_1\, \mu_2$, $\delta = \delta_1\, \delta_2$, $\boldsymbol{a} = (a_1, \cdots, a_n)$, $\odot$ is the Hadamard product, and $\boldsymbol{M}_i =$

$(\delta_i \boldsymbol{I}_n - \boldsymbol{\Lambda}^T)$ with $\boldsymbol{I}_n$ the n times n identity matrix. The quantities in Equations (3) and (4) are the characteristic response curves of a gene in dependence to the transcription factor concentrations. Later on, we use them for the NESS simulation of the gene circuit realizing the genetic logic circuit during the technology mapping. To allow for energy awareness in the technology mapping, we further establish a link between the kinetic model and its energy dissipation rate to derive an energetic response curve to the transcription factor concentration $\boldsymbol{c}$.

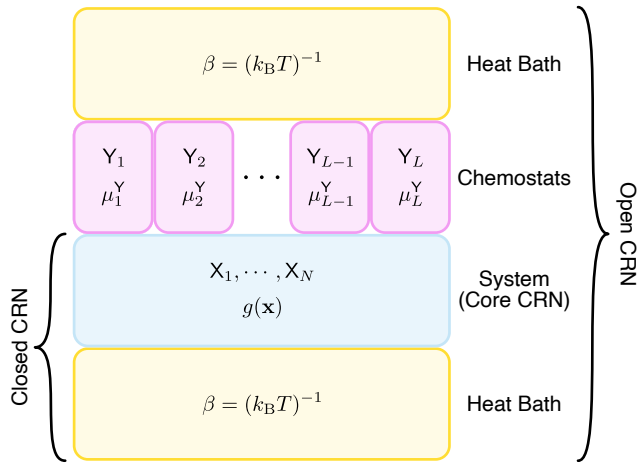


Figure 3: **Schematic description of an open CRN.** Contrasting open and closed chemical reaction networks, one observes well the difference sourced in the addition of chemostat species Y_i. The abundance of the chemostat species is kept constant, in our case by the cellular environment, and leading to a chemical potential driving the core CRN. In this work, the chemostat species refer to cellular energy carriers like ATP and the products of corresponding hydrolysis reactions. This figure is adapted from (*43*).

In stochastic thermodynamics each state $\boldsymbol{x}$ of the system is associated with its Gibbs free energy $g(\boldsymbol{x})$. Upon reaction $\mathcal{R}_m$ in forward direction the Gibbs free energy change is $\Delta g^{(m)}(\boldsymbol{x}) = g(\boldsymbol{x} + \boldsymbol{\nu}_m^X) - g(\boldsymbol{x})$, where $\boldsymbol{\nu}_m^X = [w_i^{(m)} - u_i^{(m)}]_i$ is the stoichiometric change vector. Further the system is assumed to be in contact to a heat bath with temperature T. A system consisting of a CRN and a heat bath is called a closed CRN. Following the zeroth law of thermodynamics, a closed system relaxes to equilibrium (*48*) with a mean energy dissipation rate of zero (*46*). A system in which the Gibbs free energy changes along any closed cycle in its state space add up to zero can be modelled as a closed system. For

instance, this is the case in the promoter in Figure 2**A**, if non of the transitions requires additional energy carriers. Otherwise these energy carrying species need to be accounted for in the stoichiometric equations. Furthermore, for the dynamics of RNA and protein, it is necessary to incorporate the building blocks whose recycling after degradation requires the expenditure of energy.

We thus assume that the reactions $\mathcal{R}_m$ actually include species $\mathsf{Y}_1, \ldots, \mathsf{Y}_L$, which are controlled by the cell to always have constant concentrations. Usually, these are energy carriers like ATP and its hydrolysis products like ADP and $\mathrm{PP_i}$. The coupling to these species is regarded as a coupling to a set of chemostats, which are themselves in contact with the same heat bath as the core system. Such a coupling is referred to as an open CRN, which is described with by the stoichiometric balanced equations

$$\mathcal{R}_m: \quad \sum_{i=1}^{N} u_i^{(m)} \mathsf{X}_i + \sum_{j=1}^{L} \upsilon_j^{(m)} \mathsf{Y}_j \quad \rightleftharpoons \quad \sum_{i=1}^{N} w_i^{(m)} \mathsf{X}_n + \sum_{j=1}^{L} \omega_l^{(m)} \mathsf{Y}_j. \tag{5}$$

Figure 3 illustrates the composition of closed and open CRNs. The Gibbs free energy change of the chemostat species associated with reaction $\mathcal{R}_m$ is $\Delta\mu^{(m)} = \sum_{j=1}^{L} \mu_j^{\mathsf{Y}} \left(\omega_l^{(m)} - \upsilon_j^{(m)}\right)$, where μ_j^{Y} is the chemical potential of species Y_j. Identifying $W_{\text{chem}}^{(\pm m)} = \mp \Delta\mu^{(m)}$ as the chemical work done on the core system by the chemostat upon reaction $\mathcal{R}_m$ in, respectively, forward and backward direction, we introduce the energy dissipated into the environment per reaction as the difference of chemical work and Gibbs free energy change

$$\Delta Q^{(\pm m)}(\boldsymbol{x}) = W_{\text{chem}}^{(\pm m)} - \Delta g^{(\pm m)}(\boldsymbol{x}). \tag{6}$$

The mean energy dissipation rate $\dot{Q}$ is then naturally given by

$$\dot{Q}(t) = \sum_{\boldsymbol{x}} \sum_{m=1}^{M} (\Delta Q^{(m)}(\boldsymbol{x}) \lambda_m^{+}(\boldsymbol{x}) + \Delta Q^{(-m)}(\boldsymbol{x}) \lambda_m^{-}(\boldsymbol{x})) p(\boldsymbol{X}(t) = \boldsymbol{x}). \tag{7}$$

For open CRNs the propensity functions of any microscopically reversible reaction $\mathcal{R}_m$ (i.e.

$k_m^+ > 0$ implies $k_m^- > 0$) are related to the energy changes by the thermodynamic consistency relation (*48*)

$$\ln\left(\frac{\lambda_m^+(\boldsymbol{x})}{\lambda_m^-(\boldsymbol{x})}\right) = -\beta\,\left(\Delta g^{(m)}(\boldsymbol{x}) + \Delta\mu^{(m)}\right) = \beta\,\Delta Q^{(m)}(\boldsymbol{x}), \tag{8}$$

which associates the rates of the reactions with the corresponding energy dissipation and $\beta = (k_\mathrm{B}T)^{-1}$ parameterizing the heat bath. If all reactions of an open CRN $\{\mathcal{R}_1, \cdots, \mathcal{R}_M\}$ are microscopically reversible, then we may use Equation (8) to identify the energy dissipation rate with the entropy flow into the environment

$$\frac{\mathrm{d}}{\mathrm{d}t}H^{\mathrm{env}}(t) = \beta\dot{Q}(t) = \sum_{\boldsymbol{x}}\sum_{m=1}^{M}\ln\left(\frac{\lambda_m^+(\boldsymbol{x})}{\lambda_m^-(\boldsymbol{x}+\boldsymbol{\nu}_m)}\right)J_m(\boldsymbol{x},t), \tag{9}$$

where $J_m^+(\boldsymbol{x},t) = \mathrm{p}(\boldsymbol{X}(t) = \boldsymbol{x})\,\lambda_m^+(\boldsymbol{x})$ and $J_m^-(\boldsymbol{x},t) = \mathrm{p}(\boldsymbol{X}(t) = \boldsymbol{x}+\boldsymbol{\nu}_m^X)\,\lambda_m^-(\boldsymbol{x}+\boldsymbol{\nu}_m^X)$ represent the forward and backward probability fluxes of reaction $\mathcal{R}_m$ with net flux $J_m(\boldsymbol{x},t) = J_m^+(\boldsymbol{x},t) - J_m^-(\boldsymbol{x},t)$. The entropy production rate (*46*, *49*, *50*) is the sum of the entropy change rate of the system and the entropy flow into the environment, i.e.,

$$e_p(t) = \frac{\mathrm{d}}{\mathrm{d}t}H[\boldsymbol{X}(t)] + \frac{\mathrm{d}}{\mathrm{d}t}H^{\mathrm{env}}(t) = \sum_{\boldsymbol{x}}\sum_{m=1}^{M}J_m(\boldsymbol{x},t)\,\ln\left(\frac{J_m^+(\boldsymbol{x},t)}{J_m^-(\boldsymbol{x},t)}\right), \tag{10}$$

where $H[\boldsymbol{X}(t)] = -\sum_{\boldsymbol{x}}p(\boldsymbol{X}(t) = \boldsymbol{x})\ln(p(\boldsymbol{X}(t) = \boldsymbol{x}))$. Due to the steady state assumption, $\lim_{t\to\infty}\frac{\mathrm{d}}{\mathrm{d}t}H[\boldsymbol{X}(t)] = 0$ and the mean energy dissipation rate (in units of $k_\mathrm{B}\,T$) of the NESS is given by the entropy production rate, i.e.,

$$\dot{Q} = k_\mathrm{B}T\,e_p. \tag{11}$$

Due to the central relevance of this relationship, we recommend SI Section S3 and in particular Section S3.3 to the reader, where we discuss the relationship between entropy production rate and energy dissipation rate in depth.

Following the previous, we express the energetic response curve to the transcription factor

concentrations $\boldsymbol{c}$ in terms of the NESS's overall mean energy dissipation rate $\epsilon_g(\boldsymbol{c})$. This rate is given as the sum

$$\epsilon_g(\boldsymbol{c}) = \epsilon_p(\boldsymbol{c}) + \epsilon_{tx}(\boldsymbol{c}) + \epsilon_{tl}(\boldsymbol{c}). \tag{12}$$

of the contributions of the promoter $\epsilon_p(\boldsymbol{c})$, the RNA dynamics $\epsilon_{tx}(\boldsymbol{c})$, and the protein dynamics $\epsilon_{tl}(\boldsymbol{c})$. In particular, we use the relationship between energy dissipation and entropy production to quantify the energy dissipation rate of the promoter as $\epsilon_p(\boldsymbol{c}) = k_{\mathrm{B}}T\, e_p$, since it is assumed to satisfy the microscopic reversibility requirement. Note that since the entropy production rate is calculated based only on the kinetic model, the expression is valid for a variety of different promoter structures and can be used without actual knowledge of the Gibbs free energy changes $\Delta g^{(m)}$, provided the promoter topology is accurate. As the reactions that change RNA and protein abundance are irreversible, the associated energy demands are calculated based on the knowledge of both energy changes and reaction parameters.

For deriving the NESS entropy production rate of the finite state space promoter, we make use of Schnakenberg's method (*49*) which we outline in SI Section S4. Applying this method we obtain

$$e_p = J_{o_1}\,\beta\,\tilde{\mu}_{o_1} + J_{o_2}\,\beta\,\tilde{\mu}_{o_2} + J_{o_3}\,\beta\,\tilde{\mu}_{o_3} \tag{13}$$

with

$$\begin{aligned}
\tilde{\mu}_{o_1} &= \Delta\mu_{12} + \Delta\mu_{26} + \Delta\mu_{65} + \Delta\mu_{51} \\
\tilde{\mu}_{o_2} &= \Delta\mu_{23} + \Delta\mu_{37} + \Delta\mu_{76} + \Delta\mu_{62} \\
\tilde{\mu}_{o_3} &= \Delta\mu_{34} + \Delta\mu_{48} + \Delta\mu_{87} + \Delta\mu_{73}
\end{aligned}$$

as derived in SI Section S4.1. In detail, the o_i are the cycles under consideration as presented

in Figure 2**A**, F_{o_i} is the net flux and $\tilde{\mu}_{o_i}$ the chemical potential associated to cycle o_i. Inserting this into the definition of the promoter's energy dissipation rate, we obtain

$$\epsilon_p(\boldsymbol{c}) = J_{o_1}\,\tilde{\mu}_{o_1} + J_{o_2}\,\tilde{\mu}_{o_2} + J_{o_3}\,\tilde{\mu}_{o_3}$$

with the flux (J_{o_i}) and force ($\tilde{\mu}_{o_i}$) products reassembling the well known power in electrical engineering, being the product of a voltage (force) and the current (flux) induced. Intuitively, the flux gives rise to the net rate of the reactions in the cycle while the chemical potential is the associated work done by the chemostat.

As stated previously, also the energy dissipation rates of transcription and translation can be derived in terms of their entropy production rates. We here approach the energetics from a rational perspective, while the corresponding considerations on the entropy production rates are presented in SI Section S3.4. Consider the energy required per molecule for RNA and protein first, we subsume all length dependent energy requirements for RNA synthesis and degradation in e_r and the length independent ones in $\bar{e}_r$. Taking the RNA length l_r in nucleotides into account, the energy per RNA molecule is $e_r\,l_r + \bar{e}_r$. For the protein synthesis and degradation, we analogously define the energies e_p and $\bar{e}_p$ and set l_p as the protein length in amino acids. To obtain the actual energy dissipation, we have to account for the synthesis and degradation rates, which are equal in the steady state. In particular, these rates are given by $\delta_1\ \mathsf{E}[R \mid \boldsymbol{c}]$ and $\delta_2\ \mathsf{E}[P \mid \boldsymbol{c}]$ for RNA and protein respectively, Multiplying the rate of molecule synthesis and degradation with the associated energy gives

$$\epsilon_{tx}(\boldsymbol{c}) = (e_r\,l_r + \bar{e}_r)\,\delta_1\ \mathsf{E}[R \mid \boldsymbol{c}] \qquad \epsilon_{tl}(\boldsymbol{c}) = (e_p\,l_p + \bar{e}_p)\,\delta_2\ \mathsf{E}[P \mid \boldsymbol{c}] \tag{14}$$

as the expected energy dissipation rates of the RNA and protein dynamics.

2.2 Energy Aware Technology Mapping

With the energy aware NESS model of gene expression introduced, we here present the means to incorporate it into the technology mapping. This gives rise to the Energy Aware Technology Mapping.

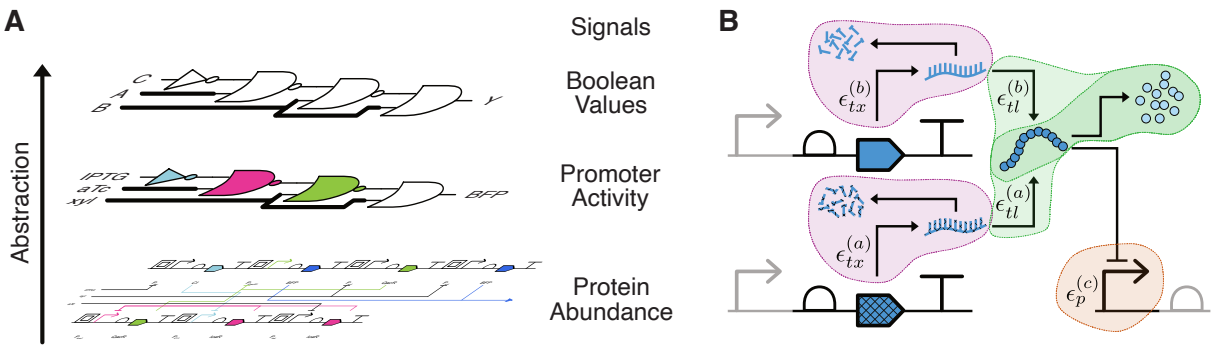


Figure 4: **From genes to logic circuits. A:** Visualization of the abstraction levels encountered in genetic design automation (GDA). The gene circuit on the bottom realizes the behaviour within the cell and is the circuit our model is applied. It relies on protein concentrations as signal carrier. Defined on top, the genetic logic circuit provides a convenient interpretation in terms of gates, familiar to engineering disciplines and the basis for GDA. The Boolean logic circuit shadows all implementational details and presents the function to realize. **B:** Overview on the energy demands of a genetic NOR gate following the implementation of (*37*). The gate includes the genes a, b, and c, with the preceding and succeeding gates greyed out. ϵ_{tx} denotes the energy dissipation rate of the RNA dynamics, ϵ_{tl} of the protein dynamics, and ϵ_p is the energy dissipation rate of the promoter.

From Genetic Logic Circuits to Genes and Back The efficient technology mapping of genetic logic circuits gets enabled by large scale *in silico* experiments. Depending on the focus, *in silico* evaluations can target different levels of abstraction as exemplified in Figure 4**A**. These abstractions not only differ with respect to the modules they are composed of, but also in the signals considered. While Boolean functions take Boolean signals into account, the genetic logic circuits we consider here the promoter activity in RPU, with protein abundance serving as signal carriers in the underlying gene circuits. With the model we present in this work, we target the gene expression level, where genes express proteins acting as transcription factors and repressing the expression level of the genes associated to their cognate promoter.

As so, this approach can reassemble the actual implementation of a genetic logic circuit within cells as presented in (*37*).

To make use of our NESS gene expression model within the genetic logic circuit centered technology mapping, we first obtain the corresponding gene circuit. In the next step, the Boolean input values are translated to the respective promoter activities, from which we then derive the corresponding inducer concentrations. The gene circuit is evaluated by applying our model in topological order to the genes. By deriving the promoter activity associated to the abundance of the reporter protein, we finally obtain the output representation at the genetic logic circuit level.

The Functional Performance of a Genetic Logic Circuit In order to complete the loop, we could apply a thresholding to the promoter activity values and obtain the associated Boolean value. Within the technology mapping, the performance of this thresholding approach is subsumed in the score S, effectively measuring the distance between activity levels representing *ON* and *OFF* states. Formally, we define the genetic logic circuit as the tuple (γ, q), where γ is the circuit's structure and q the corresponding assignment of genetic gates. We derive the score $S = S(\gamma, q)$ by evaluating the genetic logic circuit *in silico* for all Boolean input assignments $\mathcal{I}$ (with $\mid \mathcal{I} \mid = 2^{n_{\mathcal{I}}}$ for $n_{\mathcal{I}}$ inputs). The obtained circuit output values in terms of promoter activities are subsumed in $\underline{Y} = \underline{Y}(\gamma, q)$ in case they should represent an *OFF* state and $\overline{Y} = \overline{Y}(\gamma, q)$ otherwise. Identifying the elementary scoring function with s, the score $S(\gamma, q)$ is given by

$$S(\gamma, q) = \min_{\underline{y} \in \underline{Y}, \overline{y} \in \overline{Y}} s(\underline{y}, \overline{y})$$

where $\underline{y}, \overline{y}$ can be scalar values as in the case of the cello-score (*10*) or empirical distributions as for the E-score (*27*). Within Section 4.4, we extend on the definition of the E-score and it's application to empirical distributions.

The Energy Dissipation of a Genetic Logic Circuit Besides the functional characterization, our gene expression model allows for insights into the energetics of the genetic logic circuit. For this, we define $\mathcal{L}(\gamma, q)$ as the set of genetic logic gates included in the circuit (γ, q). Furthermore, we identify the genes expressing the transcription factor associated to gate $l \in \mathcal{L}(\gamma, q)$ with the index set $\mathcal{G}_C(l)$ and the genes with cognate promoters with the index set $\mathcal{G}_P(l)$. Again, the energy dissipation rate is the aggregation of the single parts contributions, resulting in

$$\epsilon_l(i) = \sum_{g' \in \mathcal{G}_P(l)} \epsilon_p^{(g')}(\boldsymbol{c}_{g'}^{(i)}) + \sum_{g \in \mathcal{G}_C(l)} \left(\epsilon_{tx}^{(g)}(\boldsymbol{c}_g^{(i)}) + \epsilon_{tl}^{(g)}(\boldsymbol{c}_g^{(i)}) \right) \tag{15}$$

where $i \in \mathcal{I}$ identifies the Boolean input condition and $\boldsymbol{c}_g^{(i)} \in \mathbb{R}_{\geq 0}^n$ denotes the cognate transcription factors' concentrations of gene g for the respective input condition. Within Figure 4**B** we present this at the example of a genetic NOR gate following the implementation of (*37*). Taking all gates $\mathcal{L}(\gamma, q)$ of the genetic logic circuit into account, the energy dissipation rate of the whole circuit for the ith input assignment is

$$E_i(\gamma, q) = \sum_{l \in \mathcal{L}(\gamma, q)} \epsilon_l(i). \tag{16}$$

We define the average and maximum expected energy dissipation rates $E(\gamma, q)$ and $E_{max}(\gamma, q)$ of the genetic logic circuit (γ, q) as

$$E(\gamma, q) = \frac{1}{\mid \mathcal{I} \mid} \sum_{i \in \mathcal{I}} E_i(\gamma, q) \qquad\qquad E_{max}(\gamma, q) = \max_{i \in \mathcal{I}} E_i(\gamma, q). \tag{17}$$

Depending on the type of application and the constraints enforced, both, E and E_{max}, are valid objective functions for the energetic optimization. In the case of stochastic evaluation, these aggregation functions extend naturally.

Technology Mapping of Genetic Logic Circuits The previous two sections have introduced different metrics for the evaluation of genetic logic circuits. It is the task of the technology mapping to design a topology from the available gate types that realizes the desired Boolean function. To this end, our tool systematically enumerates all possible variants of gate selection and topologies. For each of these possible solutions, a gate assignment has to be carried out, which is then heuristically optimized. To incorporate both metrics into the scoring of the resulting circuits, either a multi-objective or a constrained optimization can be done.

2.3 The Genetic Logic Circuit's Trade-Off between Energy and Function

With the energy aware gene expression model included into the technology mapping framework ARCTIC (*17*, *27*), we can explore the design space of genetic logic circuits with respect to both, functionality and energy effiency. For this purpose, we consider the genetic logic gates introduced with Chen et al. (*37*). In particular, we calibrate the parameters of our model to capture the response characteristics encoded in the cytometry data provided by (*37*) as described in Methods 4.3. We then use the obtained parametrizations to derive a genetic gate library for S. cerevisiae, consisting of 12 genetic logic gates utilizing nine independent transcription factors.

We start our evaluation by investigating the energy dissipation rates encountered in mapping logic circuits and their relation to the circuit's functionality. Based on the insights gained, a systematic evaluation of the circuit structure's effect on the energy of genetic logic circuits follows up. In this context, we also consider the Pareto fronts and the application of multi-objective optimization. To emphasize comparability among genetic logic circuits as well as function and energy, we introduce the energy efficiency, or briefly efficiency, $\overline{E}$ as the inverse of the energy measure (i.e. $\overline{E}(\gamma, q) = (E(\gamma, q)^{-1}$ or $\overline{E}_{max}(\gamma, q) = E_{max}(\gamma, q)^{-1})$ and apply normalization to the best scores encountered in the considered benchmark set.

Function and Energy Optima are Disjoint To assess the energy landscape of genetic logic circuits, we evaluate the technology mapping results of 33 logic circuits, each realizing a different Boolean function. The choice of circuits follows (*10*) and encompasses circuits ranging from two to seven genetic logic gates. For the purpose of this evaluation, we optimize each circuit for the two objectives function and energy and constrain the other in the respective case. While the minimum functional score to achieve is identical for all circuits, the upper limit on the energy dissipation rate during functional optimization is chosen in dependence to the gate count (see Methods 4.4).

Starting with a small circuit first, Figure 5**A** presents the function and energy optimized genetic logic circuits for 0x2F (3 gates). Despite the same structure, the different gate assignments feature significant differences in function and energy. In particular, the energy optimized version decreases the functionality by 70.7% to improve energy efficiency by 44.4% in comparison to the version optimized for function. This is a decrease in fold-change by a factor of 3.4 and a decrease in energy resource consumption by 30.8%. In addition, the disjoint optimality of function and energy indicates a trade-off between the two optimization goals considered. Increasing the number of gates does not change this impression, as we observe by considering the results for circuits 0xDF (4 gates), 0x20 (5 gates), and 0x81 (7 gates) in Figure 5**B**. However, what changes between the different circuit sizes is the absolute level of energy dissipation. For the functional optimized but also for the energy optimized circuits. While we estimate an energy dissipation rate of $29,400\ k_{\mathrm{B}}\,T\,s^{-1}$ for circuit 0x2F, this increases to $38,251\ k_{\mathrm{B}}\,T\,s^{-1}$ for circuit 0xDF, $40,901\ k_{\mathrm{B}}\,T\,s^{-1}$ for circuit 0x20, and $64,865\ k_{\mathrm{B}}\,T\,s^{-1}$ for 0x81 in the energy optimized case. With a contribution of 98.8% on average, protein synthesis and degradation account for the largest portion of energy expenditure.

Shifting the perspective to the set of all circuits, the optimization for energy increases energy efficiency up to 58.9% and on average by 37.2%.

This is possible as the efficiency of most of the circuits can be improved well (see Figure

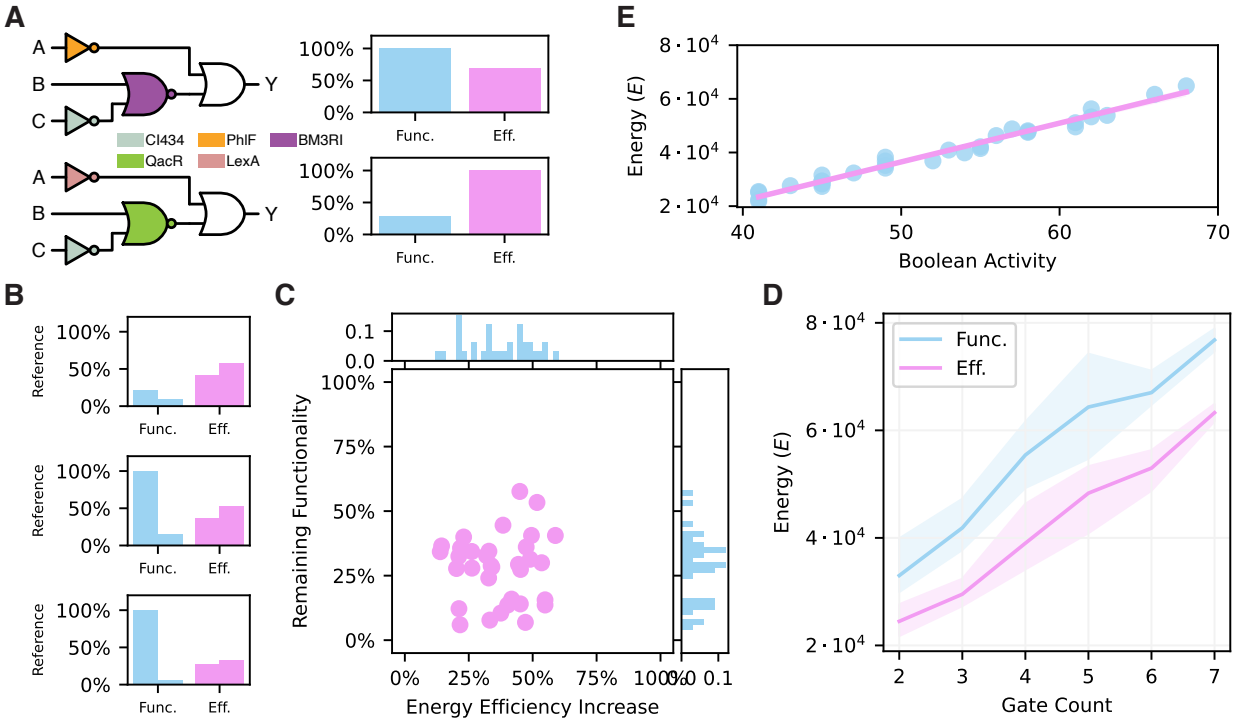


Figure 5: **Function and energy as disjoint optimization targets. A:** Circuit 0x2F (3 gates) is optimized for function (top) and energy efficiency (bottom). The bar plots on the right present the performance of the genetic logic circuits depicted on the left for both evaluation criterions. Clearly, the optima are disjoint. **B:** The optimization results of circuits 0xDF (4 gates), 0x20 (5 gates), and 0x81 (7 gates), normalized to the maximum functional score and energy efficiency observed in the benchmark. The left of a bar pair refers to the results of the functional and the right to the energetic optimization. Again, optimizing for either of the two objectives decreases the genetic logic circuits performance in relation to the other. **C:** Analysis of the cost of energy efficiency increase on the benchmark. While the optimization for energy improves energy efficiency up to 58.9%, the functionality score is often decreased significantly. **D:** Relationship between expected energy dissipation rate of the the circuit and it's gate count. The bold line is the mean and the shaded areas present the minimum and maximum intervals, with the colors indicating the optimization objective. This figure reveals a near linear relationship between energy dissipation rate and gate count. However, the optimization objective determines the offset. As so, the objective energy efficiency allows to implement genetic logic circuits of six gates with the energetic requirements of a functionally optimized four gate circuit. **E:** Boolean activity as an heuristic for the energy of a genetic logic circuit. Here compared to the energy of circuits with the objective energy efficiency.

5**C**). However, this comes at costs as often there is only a minor gap between the functionality constraint and the actual functionality score, as observable in the average remaining score of 28.5%. Figure 5**C** highlights this by visualizing the relationship between function and energy efficiency for the two cases of optimization. Again, this indicates the mutual exclusiveness of the two objectives considered and the existence of a trade-off between function and energy efficiency. Reconsidering the relation between circuit size and energy, Figure 5**E** indicates the linear relationship of these quantities for the circuits taken into account for both optimization objectives.

Guided by the relation between circuit size and energy dissipation, we aim for finding an even more precise estimate of a genetic logic circuit's energy dissipation rate. We found this in the Boolean activity of a structure, which in simplified terms counts the states which require protein expression. More formally, we consider the truthtable of the circuit structure present. For each gate, we count the number of Boolean ON states the respective input gates feature. To the inputs and outputs, special treatment applies. For the inputs, they are counted as one and the actual Boolean values are ignored as the *in vivo* implementation makes use of constitutive promoters expressing transcription factors whose abundance is regulated by the concentration of some ligand. For the outputs, we count the number of ON states as this refers to the protein level ideally representing this state. The overall Boolean activity is then obtained by summing up all the individual contributions for all Boolean input conditions possible. Figure 5**E** presents the relationship between the Boolean activity and the energy dissipation rate of a genetic logic circuits optimized with respect to energy efficiency. As the figure already indicates high correlation, the Pearson correlation coefficient between these quantities is 0.989 and for circuits optimized with respect to functionality it is still 0.94.

Reconsidering Figure 5**D** and comparing the results for energy efficiency and functionally optimized circuits among different circuit sizes, we see that the average energy of a six gate circuit optimized for energy equals or undercuts the one of a functionally optimized circuit.

This is confirmed for other circuit sizes, providing the circuits designed by Energy Aware Technology Mapping an advantage of at least a single gate.

Circuit Structure Shapes Energy Dissipation The previous results indicate a strong connection between a circuit's size and its energy dissipation rate. However, not all circuit's of a distinct size perform equally well, moving the structure of the genetic logic circuit into focus. The structural variants approach implemented in ARCTIC (*17*, *27*) allows for optimizing the structure in addition to the gate assignment within the technology mapping process. Thereby, (*27*) showcases the beneficial aspects on the function the inclusion of the structure into the optimization process has. Guided by these insights, we use the structural variants approach to evaluate the impact of the structure on a genetic logic circuit's performance with respect to both, functionality and energy efficiency.

For the evaluation of the structural variants, we consider exemplary the three Boolean functions 0xEC (3 gates), 0x02 (4 gates), and 0xE7 (5 gates), and allow for a single excess gate in comparison to the minimal structure possible. We apply constrained optimization as in the previous section but focus on the optimization with respect to energy effiency. Figure 6**A** provides a first intuition on the effect of structural variants by presenting the results of function 0xEC. The enumeration of structural variants revealed 16 different structures, which significantly differ in their functional but also energetic characteristics. The worst structure has a 54.3% higher energy dissipation rate than the best and on average the structural variants improve energy efficiency by 30.7%. Among the presented variants also the functionality varies, with a maximum improvement of 52.4% being possible. Even more interesting is the property, that some structures (i.e. 11, 13, and 15) are significantly better than others in terms of energy efficiency and functionality.

To emphasize our understanding of the relationship between energy efficiency, functionality and structure, Figure 6**B** includes for each of the considered functions a visualization showcasing the distribution of values. Structures with less gates are more energy efficient

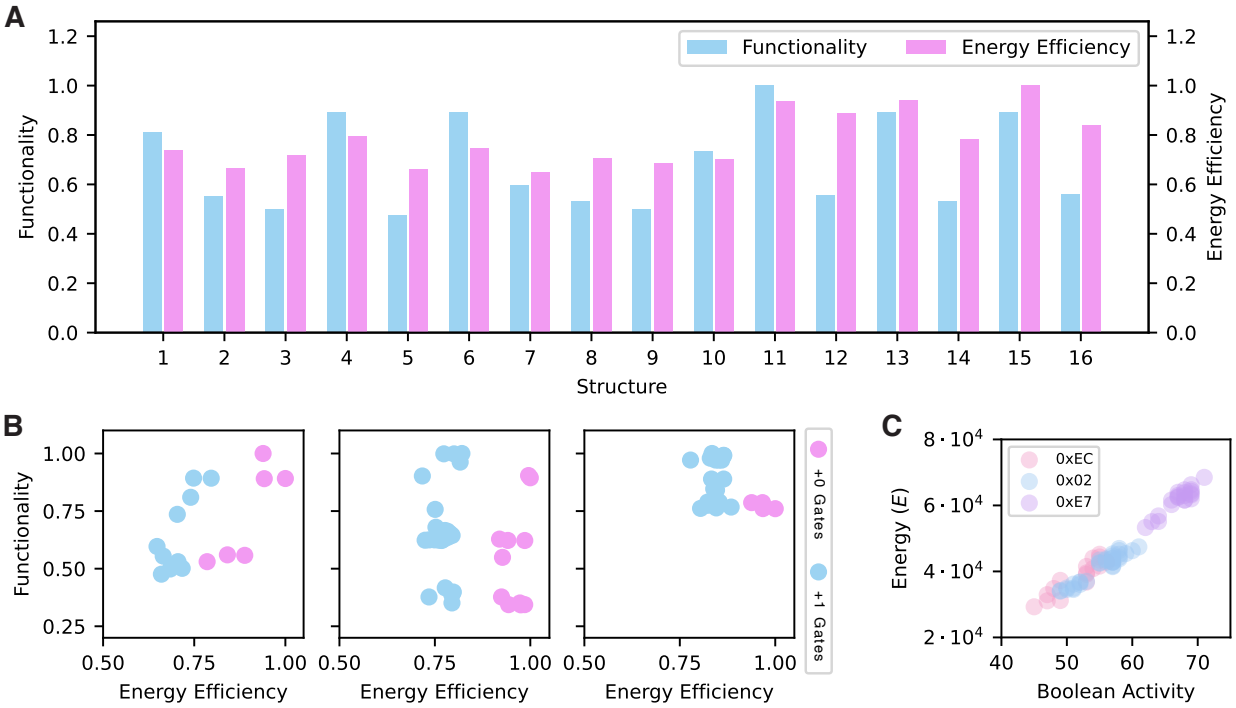


Figure 6: **The logic circuit's structure impact on energy efficiency. A:** We here present the functionality and energy efficiency of the structural variants of function $0xEC$ optimized with respect to energy efficiency. The structures feature different characteristics in both, the energy efficiency and functionality, with an average improvement of energy effiency by 30.7% when considering the best structure. Normalization refers in both cases to the best value encountered. **B:** Overview on the distribution of the variants optimized with respect to energy for the three functions 0xEC (3 gates), 0x02 (4 gates), and 0xE7 (5 gates), where the gate counts refer to the smallest structure obtained. Within the Figures, the colors code for the number of excess gates, showing that smaller structures are beneficial for energy efficiency. However, for larger circuits the consideration of excess gates proofs beneficial for functionality. The values are normalized with respect to the best values obtained for each function. **C:** Also for the structural variants, the correlation between Boolean activity and the energy dissipation is significant albeit varies among the different functions (0.927 for 0xEC, 0.96 for 0x02, and 0.934 for 0xE7).

than their larger counterparts. However, except for 0xEC, the larger circuits feature the better functional performance despite the smaller ones being rather close. Considering the energy efficiency improvements possible, these decrease with increasing circuit size as the average improvement of 0x02 is 21.4% and of 0xE7 16.1%. Besides, one observes that again some structures are more beneficial than others with respect to both objectives. The strcutues lead to superior designs.

Reconsidering the Boolean activity, we here investigate whether it provides sufficient detail to predict the energy dissipation rates of the structural variants corresponding to a single Boolean function. Presented in Figure 6**C**, Boolean activity and energy expenditure again correlate significantly, with a correlation coefficient of 0.982. However, when considering single functions, the correlation reduces to 0.927 in the worst case. Despite being worse than the others, the illustration emphasizes the reliability of this heuristic also in the case of structural variants.

Pareto Frontier Dominance among Structures Until now, we only considered the extremes of being either energetically or functional optimal. In a design approach, one would trade-off these objectives to achieve the best performance in relation to the energy spend. To this end, we analyze the Pareto front of all the structures for the Boolean function 0xF7. In particular, we perform a parameter sweep over the energy constraint for each of the 10 structures. This gives rise to the functional best genetic circuit adhering the energy constraint. The constraint is chosen to equally distribute into 20 samples over the interval of energy levels identified by an initial random sampling.

The Pareto fronts obtained showcase the exclusiveness of energetically and functional optimality by giving rise to an anti proportional characteristic. Considering exemplary the Pareto front of structure 4 (Figure 7**A**), which we present in Figure 7**C**, we observe that an increase in efficiency leads to a decrease of functionality and vice versa. In addition, we observe characteristic clustering at different levels of functional performance, likely being

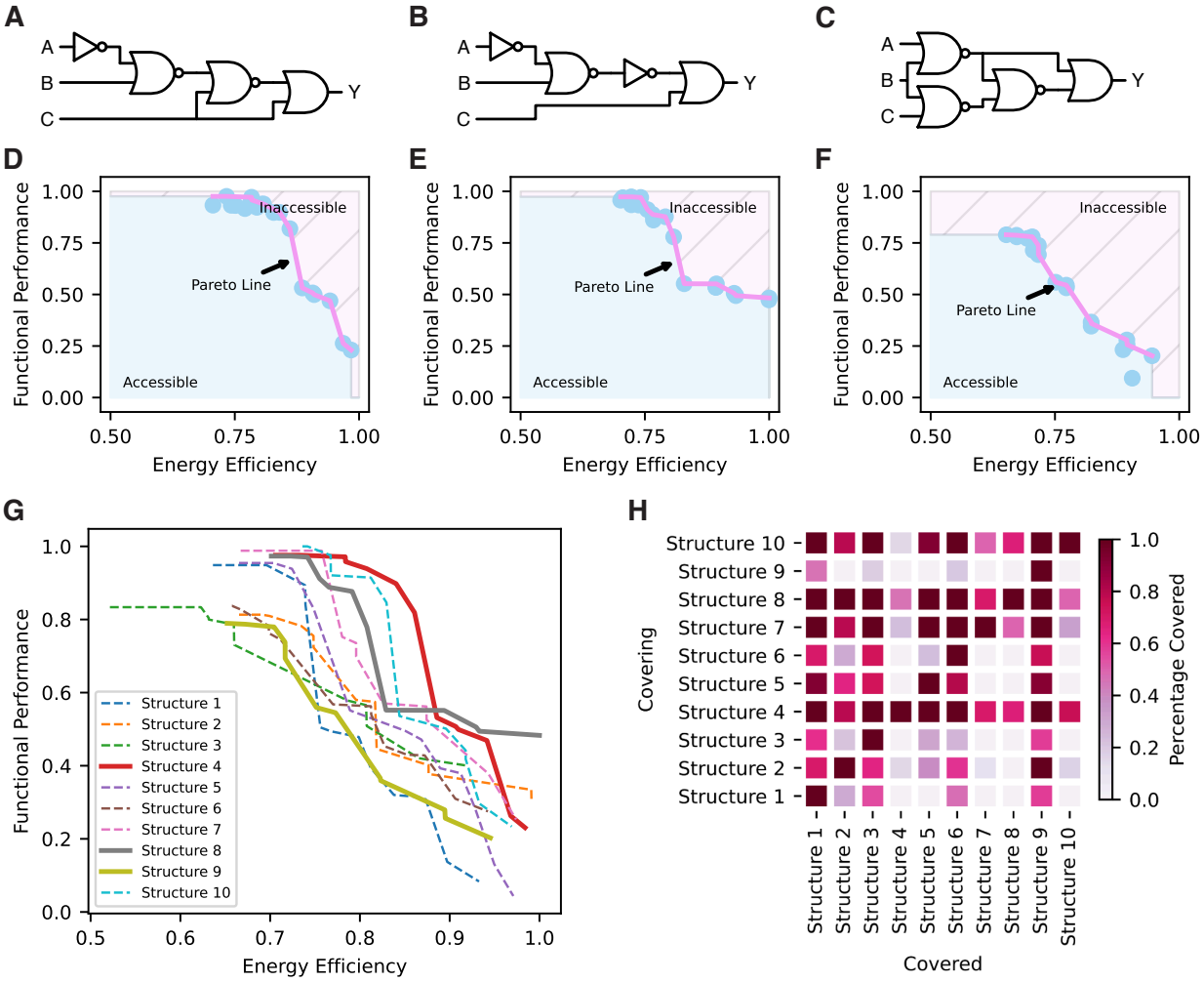


Figure 7: **The Pareto front of the boolean function 0xF7. A-C:** Three exemplary structures of the Boolean function 0xF7. In particular, Structure 4 in **A**, Structure 8 in **B** and Structure 9 in **C**. **D-F:** The Pareto fronts of the structures presented in **A-C**, where **D** presents the one of Structure 4, **E** the one of Structure 8, and **F** corresponds to Structure 9, respectively. Normalized to the same interval, the Pareto fronts (Figures **D-F**) exhibit different characteristics. **D** combines high functional performance with moderate energy requirements while **E** allows the highest energy efficiency still preserving a moderate functional performance. The Pareto front of Structure 9 exhibits a near linear relationship between function and energy efficiency, with an inferior overall performance. **G:** We here illustrate the Pareto frontiers of all the structures of 0xF7 jointly to emphasize comparability and the differences among them. Highlighting the Pareto fronts presented in Figures **D-F** by bold lines, their relationship to one another gets obvious. In addition, one observes that structure 9's Pareto front is inferior to others. Figure **H** illustrates this quantitatively, by stating the portion of Pareto front covered by another Pareto front. Thereby, the rows indicate the Structure which's Pareto front covers the respective Pareto front denoted in the column. This comparison emphasizes the dominating behaviour of structures 4, 8 and 10 over the others with respect to both optimization goals.

caused by different genetic gates at the later positions of the circuit (*17*). Despite this, the Pareto front presents a near convex behaviour and allows for a smooth trade-off between function and energy. Our evaluations of exemplary structures for the Boolean function 0x26 featuring six genetic gates each also presents the anti-proportionality but does not exhibit the observed smoothness.

Continuing with the Pareto frontiers of Structures 8 and 9 as visualized in Figures 7**E** and **F**, we observe that the general relationship persists while the exact form changes. In particular, the front in **E** appears inferior to the one in **D**, and this is surely the case for the areas with high functional performance, as Structure 4 allows for higher energy efficiency there. However, when it comes close to the limits of energy efficiency, structure 8 (7**E**) is more robust in terms of the functional score achieved. Structure 9 (7**F**) features an almost linear transition region, but is obviously inferior to both of the previous with respect and to any optimization objective and the joint optimization as well.

Inspired by the inferior performance of Structure 9 along the whole Pareto front, we compare the Pareto frontiers of all structural variants of Boolean function 0xF7 in Figure 7**G**. The Pareto frontiers are even more diverse than the ones already considered (highlighted by bold lines). Structures 4 and 8 cover almost all of the relevant optimization space, regardless of the trade-off considered. In contrast, seven out of ten structures do not allow the optimization to reach the best solution. Figure 7**H** visualizes this quantitatively by representing for each structure the portion of the Pareto points of any other structure covered. At first sight, one notices the four rows featuring a large number of dark squares. In particular, these are the rows of Structures 4, 7, 8, and 10, which cover the Pareto fronts of the superior structures either completely or almost completely and thus exhibit a sort of Pareto dominance.

Despite considering energy efficiency and function jointly, the current optimizations only optimize either of the objectives while constraining the other. Multi-objective optimization provides means to overcome this, as it allows to optimize both objectives jointly (*51*). Dealing

with the two objectives function and energy, represented by the scores S and $\overline{E}$, one approach to realize multi-objective optimization is scalarization. In the case of scalarization, the objective functions are aggregated into a single objective, to which one can apply conventional optimization methods as the simulated annealing presented in (*27*).

Due to the well behaving shape of the Pareto frontiers examined, we here suggest a linear scalarization, which is also known as the weighting method (*51*, *52*). For it's efficient application ahead of identifying the whole Pareto front, the scores need to be normalized (*51*). To this end, we introduce S_r and $\overline{E}_r$ as reference maxima of the respective scores, which one obtains by an initial evaluation. In dependence to the weight parameter $\phi \in [0, 1]$, we express the scalarized performance of the genetic logic circuit (γ, q) by $V(\gamma, q)$, which we define as

$$V(\gamma, q) = \phi \, \frac{S(\gamma, q)}{S_r} + (1 - \phi) \, \frac{\overline{E}(\gamma, q)}{\overline{E}_r}.$$

By the choice of the parameter ϕ, one implements the trade-off between the two objectives energy efficiency and function.

3 Discussion

In this work, we present an energy aware gene expression model characterizing the non-equilibrium steady state in terms of the first and second order moments and the associated energy requirements, which we relate to the model's entropy production rate.

With this model at hand, we establish **Energy Aware Technology Mapping**, the design of genetic logic circuits with respect to function and energy efficiency. The *in silico* evaluation on the basis of 33 Boolean circuits (*10*, *27*) revealed significant differences between the optimization for functionality and energy efficiency. In particular, the energy efficiency can be increased on average by 37.2%, which is accompanied by an average functional performance of 28.5% in comparison to the functionally optimized variant. To estimate the

energetic costs ahead of the *in silico* simulation, we introduce the notion of Boolean activity, achieving a correlation of 0.989 to the energy expenditure of genetic logic circuits optimized for energy. With respect to structural variants, we show that different structures of a single Boolean function vary in their energetic characteristic, however can be beneficial for function and energy efficiency. For the considered functions, the average increase in energy efficiency is 22.7%. Furthermore, we showcase the existence of Pareto dominance among structural variants, indicating the existence of beneficial characteristics for both objectives. Considering the circuit size, the evaluation shows that the Energy Aware Technology Mapping allows to implement genetic logic circuits at the energetic costs of circuits featuring at least one gate less.

The non-equilibrium gene expression model showcased in this work can account for promoter architectures varying in the number of binding sites, activation steps, and cognate transcription factors (*11*). Furthermore, the transcriptional activity is not coupled to the binding and unbinding of transcription factors but is instead modulated by these events, allowing for transcriptional bursting in the multistep activation as well as the exchange model (*11*, *22–26*). While extrinsic noise in the sense of (*53–55*) is not yet part of the model, it can be included to account for the dependence on external context factors. In contrast to the widely employed equilibrium models (*14*, *14*, *16*, *17*), the presented model can capture non-equilibrium characteristics like increased sensitivity and sharpness, which are especially relevant in the context of eukaryotes (*1*, *19*, *20*). A key aspect thereof is the dissipation of energy, being essential to all living organisms (*2*, *3*) and leading to the naming of the model as energy aware. In combination with the probabilistic NESS description and the presented relation between energy and entropy production rate, this allows for further evaluation of the trade-off between function and energy, beyond the scope of genetic logic circuits.

Despite already guided by intuition, our evaluation of the genetic logic circuits with respect to function an energy efficiency provides strong evidence that genetic logic circuits feature an inherent trade-off between functionality and energy efficiency. This matches

previous results results of (*11*, *13*) on the relation between energy expenditure and precision and sharpness. In our setting, the functional performance equals the fold change between values representing distinct Boolean states. As so, this performance can most easily be regulated by the level of protein expression, with the energy expenditure rising proportional (see Equation (14), Figure 2**C**, and Results 2.3). Also the applicability of the Boolean activity as precise heuristic for the energy dissipation is rooted in this characteristic. For a functional genetic logic circuit, the different signal levels require sufficient fold-change dynamics to allow for robust functioning. In the presence of basal expression, this is best achieved by a high expression level in the other case.

The consideration of structural variants proofs relevant also in the case of energy efficiency. While it's beneficial to consider structural variants irrespective of the optimization goal, not considering them likely leads to suboptimal solutions with respect to both, the energy efficiency and functional performance.

Considering the drop in functionality caused by the optimization with respect to energy, the circuit's performance might raise concerns on the proper functionality. However, the lower limit on the functionality score is still sufficient for robust genetic logic circuits, especially in combination with the E-score (*27*). Consequently, the current approach enables the reduction of the energetic cost of a genetic logic circuit by one to two gates. As the (energetic) resources provided by host organisms are limited, this can make the difference between the circuit being implementable *in vivo* or not.

Metabolic burden, which diverts resources from the host to synthetic constructs and impacts cellular fitness, is a critical factor to consider. Energy consumption proportional to heterologous protein expression rate is observed in many studies (*5*, *56*) as well in ours (see Figure 2**C** and Results 2.3). The possible decrease in energy showcases the possibility to reduce overall expression of heterologous proteins and in turn the burden raised by synthetic constructs. Applicable design choices are less obvious for interacting genetic circuits as considered here, highlighting the importance of our approach for coping with burden. In

that sense, the well being of the host organism ensures the desired functionality of the genetic logic circuit.

To validate the theoretical framework of our proposed Energy Aware Technology Mapping for genetic logic circuits, we outline several potential experimental evaluations. First, we propose using well-characterized genetic circuits from the literature (*37*) and compare their performance with our energy optimized versions. The simplest method to evaluate energy efficiency involves observing cell growth and viability by measuring optical density at 600 nm (OD600) and cell count (*57*). In order to measure metabolites directly linked to the cell's energetic state *in vivo*, ATP sensors could be employed (*58*, *59*). A more complex, yet insightful, method is RNA sequencing (RNA-seq). Sequencing the transcriptome of cells harboring synthetic circuits would not only reveal the transcripts associated with the circuits but also allows to check for upregulation and downregulation of genes linked to overall cell fitness (*60*). As this approach addresses the increased metabolic burden that larger circuit impose, incorporating energy dissipation while maintaining circuit functionality could enable the development of larger circuits that operate effectively *in vivo*.

4 Methods

The here presented methods present the techniques most relevant to our work. Within the referenced supplementary information, this is complemented.

4.1 Gene Circuit Simulation

The input to the gene circuit are the concentrations of the inducer molecules. While *in vivo* the inducer binds to constitutively expressed repressive transcription factors, we realize the dependency on the inducer levels by a lookup table based approach, matching a degenerate promoter model with the desired promoter activity to the provided inducer concentration. For the simulation of the gene circuit, we apply our energy aware gene expression model

one after another to the genes within the circuit. Thereby, one has to preserve topological order, meaning that a gene can only be considered when all genes expressing the cognate transcription factor have already been evaluated. To obtain the concentration of the protein encoded by a gene, we apply a log-normal closure to the moments provided by Equations (3) and (4) and draw a sample. In case multiple genes express the same protein, the protein's final concentration is the superposition of the single genes' contributions. By performing this approach many times in parallel, one obtains an empirical distribution predicting the population dynamics of cells implementing the particular genetic logic circuit. While this is a sampling based approach, one can instead propagate the mean or the median as the protein abundance to yield a point estimate. For a more formal treatment of probabilistic genetic logic circuit simulation, we refer the interested reader to (*27*).

4.2 Steady State Distribution of CTMCs and Kirchhoff's Theorem

The unique steady state distribution of an ergodic CTMC with n states can be derived by using it's propensity matrix $\mathbf{\Lambda}$, defined in row sum zero form. The first step is to identify the null-space, respectively solving for the left eigenvector $\boldsymbol{v} = [v_1 \quad \cdots \quad v_n]$ corresponding to the eigenvalue 0 of $\mathbf{\Lambda}$.

$$\boldsymbol{v}\,\mathbf{\Lambda} = \mathbf{0}$$

The steady state distribution $\boldsymbol{\pi}$ is then defined as

$$\pi_i = \frac{v_i}{v} \qquad\qquad v = \sum_{i=1}^{n} v_i.$$

Schnakenberg (*49*) and later (*61*) describe an alternative approach, called Kirchhoff's theorem -as initially described by Kirchhoff- based on the enumeration of spanning trees.

The subsequent contains a rather informal description, wherefore we refer the interested reader to (*49*) and (*61*).

Before we start with the actual method, we recall the definition of a spanning tree first. A spanning tree of an undirected graph is a tree and a subgraph, containing all the vertices of the graph. The undirected graph being a prerequisite, we derive it by identifying the vertices of the graph with the states in the state space of our CTMC. Next, the forward and backward reactions between two states are aggregated into a single undirected edge, connecting the vertices associated to the states.

We continue with enumerating all possible spanning trees of the graph. In total, there are n_t trees and we refer to the jth with T_j. Next, we introduce the function $\Pi_i(T_j)$, which maps the spanning tree T_j to the subgraph of the CTMC's state space that includes the reactions being part of the spanning tree and directed to state z_i. In particular, this is done by setting z_i as the root node and successively traversing the edges in the spanning tree T_j. Actually, each edge corresponds to two reactions, where the one pointing in the direction of the root node is preserved and the other is dropped. The result is a directed tree, consisting of the reactions that lead from any initial state to the root state z_i. With the previous, we assign the value

$$v_i = \sum_{j=1}^{n_t} \prod_{\Lambda \in \Pi_i(T_j)} \Lambda$$

to each state z_i, where the product runs over the propensities of the reactions in $\Pi_i(T_j)$. As before the steady state distribution is given by

$$\pi_i = \frac{v_i}{v} \qquad\qquad v = \sum_{i=1}^{n} v_i.$$

While in theory this method is not limited by the size of the CTMC's state space, the practical applicability depends on the complexity of the obtained graph's topology (*49*). However, this approach provides an appealing way to derive the steady state distribution

symbolically. As so, our implementation makes use of the steady state distribution derived from Kirchhoff's theorem for the promoter architectures considered, while it also provides the simpler eigenvector approach for other architectures.

4.3 Parameter Estimation

The estimation of a parameter set matching the characteristics of the part to capture is essential for the model's expressiveness within a technology mapping application. As the model gives rise to mean and variance of the response characteristic, data sources giving rise to both are ideal. One possible source is cytometry data for different inducer concentrations. As the histograms provide rich information on the distribution, the mean and variance can be derived by evaluating

$$\mu = \sum_{i=1}^{M} b_i \, \omega_i \qquad\qquad \sigma^2 = \sum_{i=1}^{M} (b_i - \mu)^2 \, \omega_i$$

where $(b_1, \cdots, b_M)$ are the bin centers and $(\omega_1, \cdots, \omega_M)$ are the frequencies in a histogram with M bins.

As model instantiation, we use a promoter architecture featuring three binding sites for a single cognate transcription factor and having only two activation steps with respect to transcriptional activity as depicted in Figure 2**AB**. Since we collapse states with equal number of transcription factors bound into a single state, we obtain a CTMC of eight states and 20 reactions transitioning between these states. In addition, we constrain all the states corresponding to the same transcriptional activation level to have the same promoter activity a_i. As the reaction rate constants for the RNA and protein dynamics are fixed in dependence to the organism, our model instantiation features 22 parameters, 20 for the reaction rate constants of the promoter's CTMC and two for the two promoter activity levels. In foregoing evaluations, this instantiation features a well balance between the complexity of the model and the capabilities of modeling response characteristics of interest, especially the steepness

encountered. However, the variance in the data is not completely explained by this, likely caused by extrinsic noise (*53–55*) not accounted for in this model.

The parameter estimation approach has to identify the values of the 22 parameters matching best to the provided data. The first step is, that we reimplement the experimental setting used to characterize the parts in (*10*, *37*) *in silico*. This setup is shown in Figure S4 and consists of two genes, whereby the genetic gate is realized by the transcription factor encoded by the first gene and the promoter on the second gene. In comparison to (*37*), we omit the gene constitutively expressing lacI, the transcription factor binding to the inducer IPTG and allowing for induction, by replacing it *in silico* with a promoter directly translating inducer levels to promoter activities. Following (*37*), the output of the gene cascade is the reporter protein. Next, the inducer levels are associated to the corresponding cytometry measurements, respectively to the associated mean and variance.

The parameter matching is achieved by minimizing the error function measuring the difference between the distribution resulting from the model setup and the data provided. As the cytometry data is given in terms of RPU, we rescale the reporter output to promoter activity levels, yielding the model's output in RPU. The error function measures the logarithmic distance between the model's mean and standard deviation to the reference data. To account for precisely matching the saturated regions, we weight the errors at the smallest and largest input conditions accordingly. In addition, a penalization is added in case the current response characteristic is not monotonous, which ensures well behaving models in domains not covered by the experimental data.

To ease optimization, the reaction rate constants for the reactions between to transcriptional activity levels are initialized to feature an inverter characteristic, while the others are chosen randomly. Furthermore, we constrain all rate constants to the range $[10^{-5}, 10^{5}]$ and the activity level of the active levels to be at least the maximum mean value encountered in the cytometry data and exceeding it by factor two at most. The inactive level is upper bounded by the smallest mean value and is allowed to undercut it by half. For each genetic

gate, the best parameter set out of ten optimizations is chosen and included in the respective promoter entry in the gate library.

4.4 ARCTIC

ARCTIC is a technology mapping framework for genetic logic circuits. It takes a combinational Boolean specification as input and it engineers all possible structural variants for this specification based on a given library of gates. It then uses different scoring methods to optimize the gate assignment of library gates to the elements of the topologies. In this way it can search for the best performing solution for a given Boolean specification. The focus of ARCTIC is the creation of *robust* implementations. To this end two different scoring methods have been implemented in previous versions of ARCTIC.

1. Takes care of the stochastic nature of cell populations. It simulates distributions for the 0-output and the 1-output. The score is then the Wasserstein metric of those two distributions. The further "away" they are from each other, the better is the robustness of the circuit. Usually, thousand randomly drawn samples are passed through the gate model to evaluate the distributions.

2. Also accounts for context effects. Crosstalk between signals is one example of such context effects. Another could be titration of transcription factors by binding to non-cognate binding sites. Those effects can degrade the performance of the circuit and thus, should be taken into account during construction.

ARCTIC offers different methods to optimize the circuit for a given Boolean specification. Exhaustive search can be used, which enumerates all possible solutions. For obvious reasons of runtime complexity, this can only be used for small examples. Branch&Bound can be used, if the gates exhibit strictly monotonic behavior. Also, the model must follow certain rules. The most versatile optimization method is Simulated Annealing. It can work with

arbitrary scoring functions and also can handle relatively large circuits. The most recent version of ARCTIC is available as open source.

Within our evaluation, we employed deterministic and sampling based simulations. The evaluation on the benchmark set as well as the structural variants made use of the deterministic approach taking the average into account. The constraint optimization used a Cello (*10*) or E-score (*27*) of 200 as lower bound on the functionality, which reassembles a 200 fold-change on the mean values. During optimization with respect to function, the upper bound on the energy expenditure was set in dependence to the circuits size, allowing for 20.000 $k_\mathrm{B}\,T\,s^{-1}$ per gate. In the Pareto evaluation, we employ the sampling based approach and consider 1000 samples for each output distribution.

E-Score Circuit-scoring in ARCTIC is done using the E-score (*27*), a generalization of Cello's circuit score (*10*) to the case where circuit outputs are random variables. The E-score is calculated as a modified 1-Wasserstein distance between empirical distributions of the concentrations or copy-numbers of the output chemical species. Consider two sets of N samples $\mathcal{X}_0 \equiv \{x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(N)}\}$ and $\mathcal{X}_1 \equiv \{x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(N)}\}$ that represent two empirical circuit output distributions. The set $\mathcal{X}_0$ consists of those samples corresponding to an anticipated circuit output of logical zero (OFF) and $\mathcal{X}_1$ consists of those samples corresponding to an anticipated circuit output of logical one (ON). We then calculate the modified 1-Wasserstein distance $D(\mathcal{X}_1, \mathcal{X}_0)$ by

$$D(\mathcal{X}_0, \mathcal{X}_1) = \exp\left[\frac{1}{N}\left(\sum_{n=1}^{\lfloor \frac{N}{2} \rfloor} \log(x_1^{(n)}) - \log(m_0) + \sum_{n=\lfloor \frac{N}{2} \rfloor+1}^{N} \log(m_1) - \log(x_0^{(n)})\right)\right], \quad (18)$$

where m_0 and m_1 are the medians of $\mathcal{X}_0$ and $\mathcal{X}_1$ respectively. Consider now having several sets $\mathcal{X}_0^1, \mathcal{X}_0^2, \ldots, \mathcal{X}_0^K$ representing logical OFF states and $\mathcal{X}_1^1, \mathcal{X}_1^2, \ldots, \mathcal{X}_1^L$ logical ON states, where each individual set corresponds to one of $K + L$ different input configurations of the circuit, i.e. different combinations of input inducer concentrations. The circuit's E-score is

then obtained by

$$S = \min_{k \in \{1, \ldots, K\}, l \in \{1, \ldots, L\}} D(\mathcal{X}_0^k, \mathcal{X}_1^l). \tag{19}$$

It is shown in (*27*) that S simplifies to Cello's circuit score when only the medians m_0^k and m_1^l of $\mathcal{X}_0^k$ and $\mathcal{X}_1^l$ are considered. Note, that the exponentiation in eq. (18) can be omitted in finding the minimum in eq. (19) and then be applied afterwards to obtain the score S.

5 Code Availability

The model is implemented as part of the technology mapping framework ARCTIC, which is available at https://www.rs.tu-darmstadt.de/ARCTIC.

6 Associated Content

6.1 Supplementary Information

Derivations of moment equations, the chemical reaction network of the model presented, background on chemical reaction networks and other techniques used.

7 Author Contributions

H.K.and C.H. provided the research idea and contributed methodology. E.K. derived the model and implemented the model and the Energy Aware Technology Mapping in ARCTIC. M.G. and E.K. performed mathematical analysis of stochastic thermodynamics. T.S. provided the interface for ARCTIC and performed evaluation. J.M. and M.M. contributed to the biological background. N.E. contributed to ARCTIC and the E-score. All authors contributed to the writing of the paper.

8 Acknowledgments

The authors acknowledge all people in the Self-Organizing Systems lab for the valuable discussions. We specially want to thank Eike Mentzendorff for advice on multi-objective optimization.

For the use of SBOLCanvas, we acknowledge Terry et al. (*62*), Quinn et al. (*29*), and the SBOL community (*28*).

9 Competing Interests

The authors declare no competing interests.

References

1. Zoller, B., Gregor, T., and Tkačik, G. (2022) Eukaryotic gene regulation at equilibrium, or non? *Current Opinion in Systems Biology 31*, 100435.

2. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Molecular Biology of the Cell*, 6th ed.; Garland Science: New York, NY, 2017.

3. Ahsendorf, T., Wong, F., Eils, R., and Gunawardena, J. (2014) A framework for modelling gene regulation which accommodates non-equilibrium mechanisms. *BMC Biology 12*, 1–23.

4. Broach, J. R. (2012) Nutritional Control of Growth and Development in Yeast. *Genetics 192*, 73–105.

5. Kastberg, L. L. B., Ard, R., Jensen, M. K., and Workman, C. T. (2022) Burden Imposed by Heterologous Protein Production in Two Major Industrial Yeast Cell Factories: Identifying Sources and Mitigation Strategies. *Frontiers in Fungal Biology 3*, 827704.

6. Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. A. (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in Biotechnology 34*, 652–664.

7. Nikolados, E.-M., Weiße, A. Y., Ceroni, F., and Oyarzún, D. A. (2019) Growth Defects and Loss-of-Function in Synthetic Gene Circuits. *ACS Synthetic Biology 8*, 1231–1240.

8. Ceroni, F., Algar, R., Stan, G.-B., and Ellis, T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods 12*, 415–418.

9. Weiße, A. Y., Oyarzún, D. A., Danos, V., and Swain, P. S. (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences 112*, E1038–E1047.

10. Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., and Voigt, C. A. (2016) Genetic circuit design automation. *Science 352*, aac7341.

11. Lammers, N. C., Flamholz, A. I., and Garcia, H. G. (2023) Competing constraints shape the nonequilibrium limits of cellular decision-making. *Proceedings of the National Academy of Sciences 120*, e2211203120.

12. Barato, A. C., and Seifert, U. (2015) Thermodynamic Uncertainty Relation for Biomolecular Processes. *Physical Review Letters 114*, 158101.

13. Seifert, U. (2018) Stochastic thermodynamics: From principles to the cost of precision. *Physica A: Statistical Mechanics and its Applications 504*, 176–191.

14. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005) Transcriptional regulation by the numbers: models . *Current Opinion in Genetics & Development 15*, 116–124.

15. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005) Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development 15*, 125–135.

16. Morrison, M., Razo-Mejia, M., and Phillips, R. (2021) Reconciling kinetic and thermodynamic models of bacterial transcription. *PLOS Computational Biology 17*, e1008572.

17. Engelmann, N., Schwarz, T., Kubaczka, E., Hochberger, C., and Koeppl, H. (2023) Context-Aware Technology Mapping in Genetic Design Automation. *ACS Synthetic Biology 12*, 446–459.

18. Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016) Information Integration and Energy Expenditure in Gene Regulation. *Cell 166*, 234–244.

19. Eck, E., Liu, J., Kazemzadeh-Atoufi, M., Ghoreishi, S., Blythe, S. A., and Garcia, H. G. (2020) Quantitative dissection of transcription in development yields evidence for transcription-factor-driven chromatin accessibility. *eLife 9*, e56429.

20. Grah, R., Zoller, B., and Tkačik, G. (2020) Nonequilibrium models of optimal enhancer function. *Proceedings of the National Academy of Sciences 117*, 31614–31622.

21. Hopfield, J. J. (1974) Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proceedings of the National Academy of Sciences 71*, 4135–4139.

22. Meeussen, J. V., and Lenstra, T. L. (2024) Time will tell: comparing timescales to gain insight into transcriptional bursting. *Trends in Genetics*

23. Pomp, W., Meeussen, J. V., and Lenstra, T. L. (2024) Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation . *Molecular Cell 84*, 1036–1048.

24. Nicolas, D., Phillips, N. E., and Naef, F. (2017) What shapes eukaryotic transcriptional bursting? *Molecular BioSystems 13*, 1280–1290.

25. Lammers, N. C., Kim, Y. J., Zhao, J., and Garcia, H. G. (2020) A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting. *Current Opinion in Cell Biology 67*, 147–157, Differentiation and disease.

26. Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. (2005) Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell 123*, 1025–1036.

27. Schladt, T., Engelmann, N., Kubaczka, E., Hochberger, C., and Koeppl, H. (2021) Automated Design of Robust Genetic Circuits: Structural Variants and Parameter Uncertainty. *ACS Synthetic Biology 10*, 3316–3329.

28. Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Roehner, N., Wilson, M. L., Adam, L., Anderson, J. C., and others (2014) The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology 32*, 545–550.

29. Quinn, J. Y., Cox III, R. S., Adler, A., Beal, J., Bhatia, S., Cai, Y., Chen, J., Clancy, K., Galdzicki, M., Hillson, N. J., and others (2015) SBOL Visual: A Graphical Language for Genetic Designs. *PLOS Biology 13*, e1002310.

30. Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J. D., and Hillson, N. J. (2013) PaR-PaR Laboratory Automation Platform. *ACS Synthetic Biology 2*, 216–222.

31. Salis, H. M. *Methods in enzymology*; Elsevier, 2011; Vol. 498; pp 19–42.

32. LaFleur, T. L., Hossain, A., and Salis, H. M. (2022) Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nature Communications 13*, 5159.

33. Myers, C. J., Barker, N., Jones, K., Kuwahara, H., Madsen, C., and Nguyen, N.-P. D. (2009) iBioSim: a tool for the analysis and design of genetic circuits. *Bioinformatics 25*, 2848–2849.

34. Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., Awan, A. R., Gilbert, C., Stan, G.-B., and Ellis, T. (2018) Burden-driven feedback control of gene expression. *Nature Methods 15*, 387–393.

35. McBride, C. D., Grunberg, T. W., and Del Vecchio, D. (2021) Design of genetic circuits that are robust to resource competition. *Current Opinion in Systems Biology 28*, 100357.

36. Sechkar, K., Steel, H., Perrino, G., and Stan, G.-B. (2024) A coarse-grained bacterial cell model for resource-aware analysis and design of synthetic gene circuits. *Nature Communications 15*, 1981.

37. Chen, Y., Zhang, S., Young, E. M., Jones, T. S., Densmore, D., and Voigt, C. A. (2020) Genetic circuit design automation for yeast. *Nature Microbiology 5*, 1349–1360.

38. Jones, T. S., Oliveira, S. M. D., Myers, C. J., Voigt, C. A., and Densmore, D. (2022) Genetic circuit design automation with Cello 2.0. *Nature Protocols 17*, 1097–1113.

39. Li, C., Cesbron, F., Oehler, M., Brunner, M., and Höfer, T. (2018) Frequency modulation of transcriptional bursting enables sensitive and rapid gene regulation. *Cell Systems 6*, 409–423.

40. Otero-Muras, I., and Banga, R., Julio Automated Design Framework for Synthetic Biology Exploiting Pareto Optimality. *ACS Synthetic Biology 6*, 1180–1193.

41. Watanabe, L., Nguyen, T., Zhang, M., Zundel, Z., Zhang, Z., Madsen, C., Roehner, N., and Myers, C. (2018) iBioSim 3: A Tool for Model-Based Genetic Circuit Design. *ACS Synthetic Biology 8*, 1560–1563.

42. Baig, H., and Madsen, J. A Top-down Approach to Genetic Circuit Synthesis and Optimized Technology Mapping. Proceedings of 9th International Workshop on Bio-Design Automation, IWBDA 2017. 2013.

43. Schmiedl, T., and Seifert, U. (2007) Stochastic thermodynamics of chemical reaction networks. *The Journal of Chemical Physics 126*.

44. Zhang, J., and Zhou, T. (2014) Promoter-mediated Transcriptional Dynamics. *Biophysical Journal 106*, 479–488.

45. Leppek, K., Das, R., and Barna, M. (2018) Functional 5 UTR mRNA structures in eukaryotic translation regulation and how to find them . *Nature Reviews Molecular Cell Biology 19*, 158–174.

46. Qian, H., and Ge, H. *Stochastic Chemical Reaction Systems in Biology*; Springer International Publishing, 2021.

47. Hasenauer, J., Wolf, V., Kazeroonian, A., and Theis, F. J. (2014) Method of conditional moments (MCM) for the Chemical Master Equation. *Journal of Mathematical Biology 69*, 687–735.

48. Rao, R., and Esposito, M. (2018) Conservation laws and work fluctuation relations in chemical reaction networks. *The Journal of Chemical Physics 149*.

49. Schnakenberg, J. (1976) Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern Physics 48*, 571–585.

50. Lebowitz, J. L., and Spohn, H. (1999) A Gallavotti–Cohen-Type Symmetry in the Large Deviation Functional for Stochastic Dynamics. *Journal of Statistical Physics 95*, 333–365.

51. Miettinen, K. *Nonlinear Multiobjective Optimization*; Springer Science & Business Media, 1999; Vol. 12.

52. Gass, S., and Saaty, T. (1955) The computational algorithm for the parametric objective function. *Naval Research Logistics Quarterly 2*, 39–45.

53. Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., and Koeppl, H. (2012) Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences 109*, 8340–8345.

54. Zechner, C., Unger, M., Pelet, S., Peter, M., and Koeppl, H. (2014) Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods 11*, 197–202.

55. Albayrak, C., Jordi, C. A., Zechner, C., Lin, J., Bichsel, C. A., Khammash, M., and Tay, S. (2016) Digital Quantification of Proteins and mRNA in Single Mammalian Cells . *Molecular Cell 61*, 914–924.

56. Glick, B. R. (1995) Metabolic load and heterologous gene expression . *Biotechnology Advances 13*, 247–261.

57. Klumpp, S., Zhang, Z., and Hwa, T. (2009) Growth-rate dependent global effects on gene expression in bacteria. *Cell 139*, 1366–1375.

58. Deng, Y., Beahm, D. R., Ionov, S., and Sarpeshkar, R. (2021) Measuring and modeling energy and power consumption in living microbial cells with a synthetic ATP reporter. *BMC Biology 19*, 101.

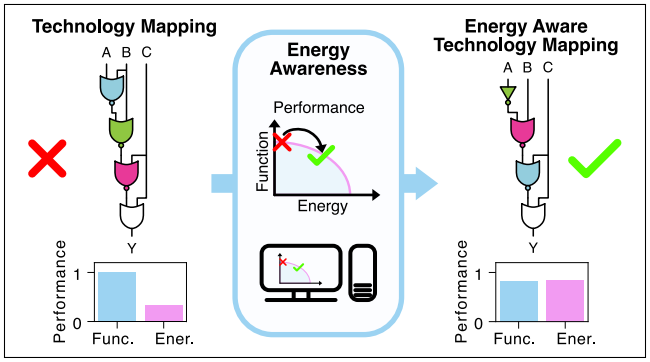
59. Botman, D., Van Heerden, J. H., and Teusink, B. (2020) An Improved ATP FRET Sensor For Yeast Shows Heterogeneity During Nutrient Transitions. *ACS Sensors 5*, 814–822.

60. Gorochowski, T. E., Espah Borujeni, A., Park, Y., Nielsen, A. A., Zhang, J., Der, B. S., Gordon, D. B., and Voigt, C. A. (2017) Genetic circuit characterization and debugging using RNA-seq. *Molecular Systems Biology 13*, 952.

61. Andrieux, D., and Gaspard, P. (2007) Fluctuation Theorem for Currents and Schnakenberg Network Theory. *Journal of Statistical Physics 127*, 107–131.

62. Terry, L., Earl, J., Thayer, S., Bridge, S., and Myers, C. J. (2021) SBOLCanvas: A Visual Editor for Genetic Designs. *ACS Synthetic Biology 10*, 1792–1796.

TOC Graphic


Technology Mapping
A B C
Y
Performance
1
0
Func.
Ener.
Energy Awareness
Performance
Function
Energy
Energy Aware Technology Mapping
A B C
Y
Performance
1
0
Func.
Ener.