Incoherent Feed-Forward Loop Models and Analysis

Tinh Son, Prof. Patel

March, 2020

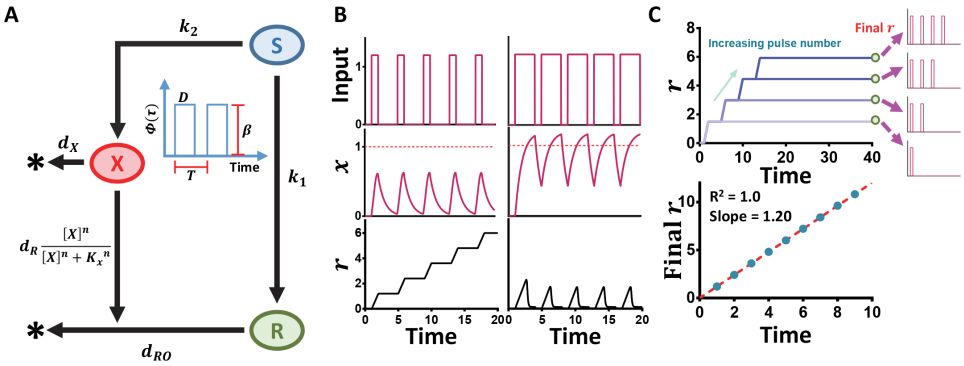
1 Introduction

The ultimate goal of synthetic biology is to one day have the ability to interface with natural system. The fundamental towards this step is to gain understanding of how biological circuit interprets or filters inputs under noisy environments. Under noise, cells require robust mechanisms to respond to both amplitude and frequency based inputs. As evidence from calcium signaling to p53 tumor suppressor cycle, oscillatory (frequency) inputs play major role in determining cell pathways [1][2]. A common motif for cell to interpret oscillatory signals or fold changes rather than concentration is the incoherent feed-forward loop (IFFL) [3].

Using synthetic biology to mimic the behaviors of natural systems is a challenging task without mathematical representation and an understanding of modeling parameters. Especially when natural biological schemes are complex and stochastic, accurately modeling circuit motifs requires correct initial assumptions that are later built upon through experimentation and revision. The aspect of why real-life circuits behave a certain way could also be explained through comparison to its alternatives. This investigation, as part of the W&M iGEM 2018 project, aims to demonstrate the process of deriving a viable model that describes the behaviors of Type I Incoherent Feed-Forward Loop (IFFL), the comparison to a direct feed-forward (or naive) loop through various novel analysis techniques with respect to dynamical inputs, and possible explanation of the motif's innocuous appearance in biological processes.

1.1 An Abstract IFFL

Zhang and his colleagues published a paper describing the generic model of IFFL [4]. Here, the input structure is described as periodic square-wave function of varying durations. This structure induces the production of an intermediate protease and a reporter, of which the reporter is quickly degraded as the protease reaches its half-activation threshold at long duration time-steps. This abstract model emphasizes IFFL's ability to distinguish distinct temporal structure of the input through identification of the reporter's steady-state behavior.



$$\frac{dr}{d\tau} = \beta\Phi(\tau) - (\gamma_R \frac{x^n}{x^n + 1} + \gamma_o)r,$$

$$\frac{dx}{d\tau} = \beta\Phi(\tau) - x,$$

Figure.1: From Zhang et al. paper, reporter R is quickly degraded by protease X through Hill's kinetics. One implication of this interaction would be the tuning of Hill's coefficient. **A.** depicts the general scheme of IFFL with input S, protease X, and reporter R. **B.** describes the output behavior of reporter R given two distinct oscillatory inputs: Longer input duration allows expression of protease X to reach its half-activation threshold and quickly degrades R, which explains the pulse-like behavior. **C.** illustates IFFL's ability to directly translate its stair-case output to exact duration of on time for the input (counting ability). Same concept can be applied to the reporter's pulse-like behavior, which describes a completely different on time duration.

Zhang highlights characteristics of reporter's output based on parameters such as production rate β and endogenous degradation rate γ_0. As part of iGEM 2018 project, this general abstract model serves as a basis for experimental evaluation as it lays down necessary components that provides

desired outputs, such as the interaction of Hill's kinetics, for the derivation of system of ordinary differential equations. Zhang derived the system of ODE above from the following system, where X and R represents the protease and reporter respectively:

$$\frac{d[R]}{dt} = k_1\,\Phi(t) - d_R\,\frac{[X]^n}{[X]^n + K_x^n}\,[R] - d_{Ro}[R]$$

$$\frac{d[X]}{dt} = k_2\Phi(t) - d_x[X]$$

Analytically solving system yields the following solutions:

$$X = \frac{k_2 - e^{-d_X(t+C)}}{d_X}$$

$$R = \frac{k_1(-e^{-(C+d_X)t}+k_2)^{-n}((d_X K_X)^n + (-e^{-(c+d_X)t}+k_2)^n)}{d_R} + e^{\frac{-d_R t(-e^{-(c+d_X)t}+k_2)^n}{(d_x K_X)^n + (k_2 - e^{-(c+d_X)t})^n}}$$

The rest of Zhang's paper analyzes the parameter conditions, such as production rate and duration, on the quality of reporter output in terms of pulse counting ability. Though this is useful, this investigation will focus solely on the holistic temporal distinguishing capability of IFFL through other means.

2 Modeling

2.1 Chemical-Inducible Model

As part of W&M iGEM 2018 project, this specific IFFL model is designed for induction, where IPTG is the activating input that inhibits the inhibition of LacI to pLac, thus clearing the production pathway for protease mf-Lon and reporter m-Scarlet.

$$I = \text{Modified Square-wave function (input duration)}$$
$$\dot{M} = \frac{\beta_M}{1+L/k_p/(1+(I/k_I)^2)} - \alpha_M[M]$$
$$\dot{S} = \frac{\beta_S}{1+L/k_p/(1+(I/k_I)^2)} - \alpha_{MS}\frac{[M]^2}{[M]^2+1}[S] - \alpha_S[S]$$

The denominator aims to highlight the minuscule leakage rate when no inducer is presented ($\frac{\beta_M}{1+L/k_p}$). The term $1 + (I/k_I)^2$ gives the concentration of unbound LacI through cooperative reaction. Based on the abstract model, Hill kinetics is kept for m-Scarlet to induce its pulse-like outputs

Symbols	Description
β_M	Production rate of mf-Lon
α_M	Degradation rate of mf-Lon
β_S	Production rate of m-Scarlet
α_S	Degradation rate of m-Scarlet
α_{MS}	Degradation rate of m-Scarlet by mf-Lon
L	LacI concentration
I	IPTG concentration
k_p	Dissociation constant between LacI and pLac
k_I	Dissociation constant between LacI and IPTG

when the duration of IPTG input is long. Certain parameters such as k_p and k_I are constants found in literature, which facilitates the model fitting process later on. Testing of this system of ODE was done through numerical calculation using Python's scipy library. The default method is ODE45.

2.2 Model Fitting with Experimental Results, Based on W&M iGEM 2018 Spin Down Test

The overarching goal for this specific model to find out whether it is possible to fit model parameters with results from the Inducer Spin Down test with WM17_411 + 393.

Fluorescence data were normalized to facilitate model fitting. We began by using Markov Chain Monte Carlo (MCMC) to find parameter range that would best fit with the model. Python's emcee library was used for this process, and Gaussian noise were added to the calculated data to further simulate real-life behaviors. Likelihood function was numerically written and optimized for calculation of the posterior with 1000 burn-steps. The result yields this following graph where parameters were best fitted. Parameter correlations were almost non-existence based on the relative distribution of parameter values.

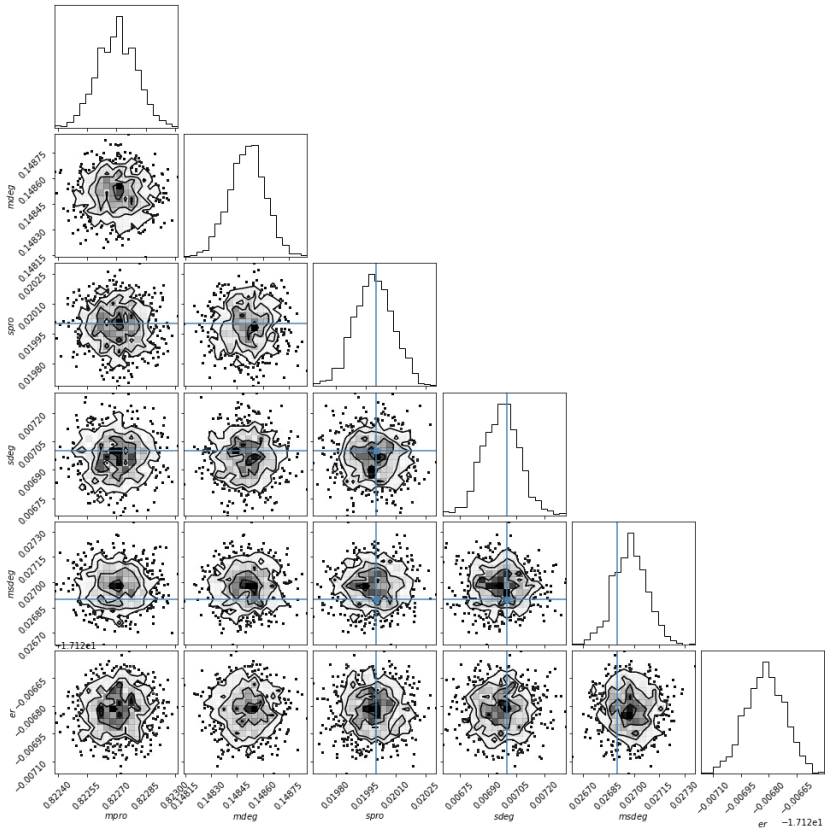


Figure. 2: Triangle graph depicting distribution density of parameter values and relative correlation to other parameters.

MCMC was not enough for accurate fitting. This is due to the lack of experimental data, where a total of 18 data points were used to estimate values for 4 different parameters. The next step was to hard code parameter values with respect to Period and Duration of input. Below is the parameter values and fitted curves.

Symbols	Value	Units	Source
β_M	0.16	$\frac{nM}{min}$	MCMC
α_M	0.13	$\frac{min^{-1}}{nM}$	Hard-coded
β_S	2.21	$\frac{nM}{min}$	Hard-coded
α_S	0.01	$\frac{min^{-1}}{nM}$	Hard-coded
α_{MS}	0.26	$\frac{min^{-1}}{nM}$	MCMC
L	0.5	nM	Hard-coded
I	100000	nM	Conversion from experiment
k_p	0.001	nM	[5]
k_I	1000	nM	[5]

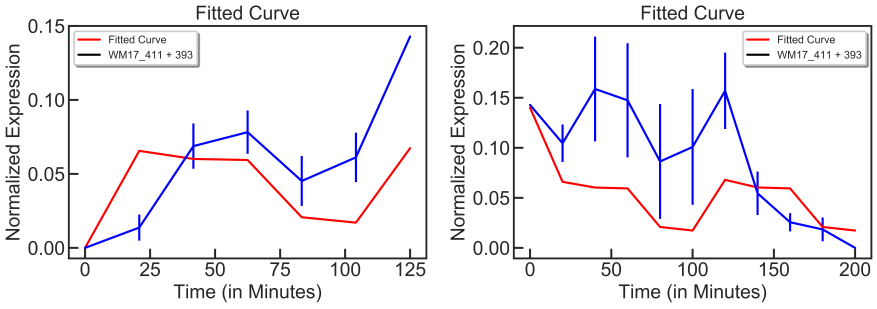


Figure. 3: Fitted curves using the above parameter values (Staircase on the left, pulse on the right)

Fitted curved from these graph are not run with noisy inputs, which could explain the differences. Furthermore, the nature of the experiments allows very high noise in expression read-outs: manual induction and spin-out of chemicals are very prone to errors, not to mention the tedious wait time for each step following induction could pave pathway to high inaccuracy. This chemical model, however, could always be improved upon to account for delay-time induction, although no necessary component would be included if a highly consistent machine were to perform the experiment (microfluidic devices).

2.3 Heat-Inducible Model

Due to the feasibility of performing experiments through chemical induction, a new heat-induced IFFL model is developed.

$$\dot{C_I} = \begin{cases} -\alpha & \text{when } T = 37°C \\ \beta_{C_I} - d_{C_I}[C_I] & \text{else.} \end{cases} \quad [C_I] = \begin{cases} 0 & \text{when } T = 37°C \\ \frac{\beta_{C_I}}{d_{C_I}} - \frac{\beta_1}{d_{C_I}} e^{-d_{C_I} t} & \text{else.} \end{cases}$$

$$\dot{M} = \frac{\beta_M}{1+[C_I]/k_d} - d_M[M]$$
$$\dot{S} = \frac{\beta_S}{1+[C_I]/k_d} - d_{MS}\frac{[M]^n}{[M]^n+1}[S] - d_S[S]$$
$$\text{For Naive System: } \dot{S} = \frac{\beta_S}{1+[C_I]/k_d} - d_S[S]$$

Symbols	Description
α	Degradation rate of CI at $37°C$
β_{C_I}	Production rate of CI
d_{C_I}	Degradation rate of CI
β_M	Production rate of mf-Lon
d_M	Degradation rate of mf-Lon
β_S	Production rate of m-Scarlet
d_S	Degradation rate of m-Scarlet
d_{MS}	Degradation rate of m-Scarlet by mf-Lon
k_d	Dissociation constant between CI and promoter

In this system, the promoter inhibitor protein CI is degraded by heat at high temperature. Model fitting was not performed since there was no data from experiment. However, comparison between this specific IFFL model and its naive model was extensively analyzed. These comparison aspects are IFFL's temporal structure distinguishing abilities and its response to noise. For the comparison, the design is such that the term α will completely degrade the concentration of CI when the temperature is switched to $37°C$.

2.4 Methods and Analysis

2.4.1 Time-input Distinguishing Ability

The ability to detect small differences in signal input is not only appealing but paramount to signal processing. For this aspect, the following numerical comparison is implemented:

With fixed parameter values, concentration expression of the reporter using all possible combinations of off-on ratio for the input are calculated. Next, the differences of these expressions are represented as function of euclidean distance. Finally, output are delineated by the vector difference between the expression of IFFL compared to naive, taken at fixed off time and different on time, with positive difference (red) being the time ratios where IFFL is better at distinguishing the input signal from one another.

Example: Take off ratio of 1, and on ratio of 1 and 3. The concentration expression is calculate with ratio of 1:1 and 1:3 for IFFL and naive. Next, the Euclidean distance of the differences between the concentration function with respect to these two on times is calculated, with the result represented as one value. Finally, the vector value from IFFL is subtracted from the one in naive to yield a final value indicative of model's fidelity in information processing.

The heat maps below are results from this comparison:

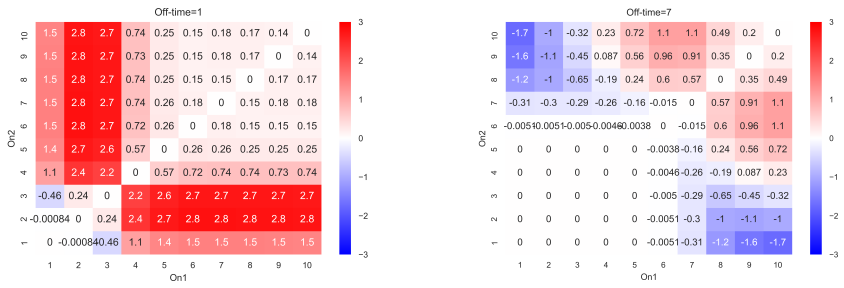


Figure. 4: Red squares are indicative of where IFFL outperforms naive at signal processing, and shading represents relatively how much better. Notice the symmetry of the graph through the diagonal. Based on the right heat map, IFFL performs worse as off time increases since the relative ratio of on time to period is very short. However, when on time ratio is close or greater than off time, IFFL once again performs better than the naive model.

2.4.2 Noise

With IFFL's ability to detect small changes in temporal structure from the input, one would expect that the system would be susceptible to noise. We

performed the following numerical testing for both IFFL and naive model and compare the variance as direct translation of output's noisiness. Here, off-on ratio refers to the time ratio when the heat is turned on (CI degrades completely) or off, analogous to Period being off + on and Duration being the on time. Gaussian noise with average value 0 and standard deviation of 0.1 was used for the concentration of CI:

Start with fixed value for off-on ratio, calculate all concentration expression of m-Scarlet within a specific range of m-Scarlet production rate ($0.35 \leq \beta_S \leq 1$). Variance is then taken for that specific off-on ratio, then on ratio is incremented (max 5 on), then off ratio (max 5) is incremented as on ratio resets to 1. Each variance for respective off-on ratio is the average variance of concentration expression as function for varying parameter value range of m-Scarlet production rate.

The graph below illustrates the result of this test and data can be found in the Supplementary Section:

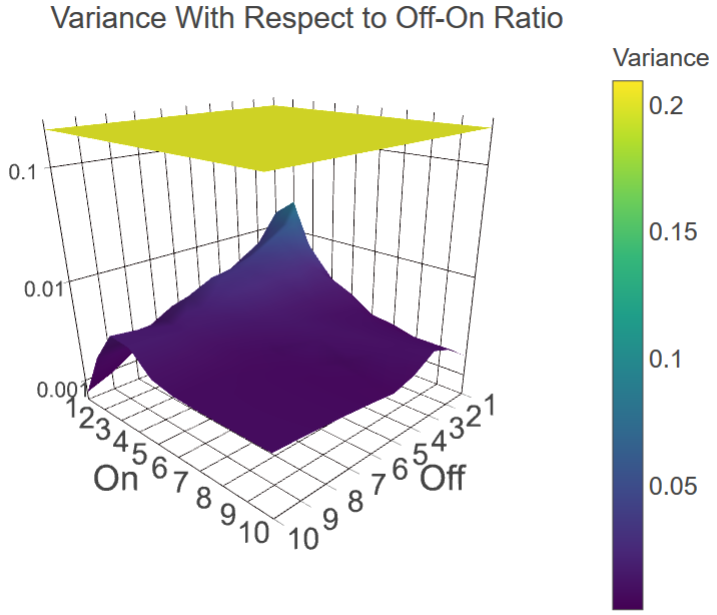


Figure. 5: Variance as direct translation to noise.

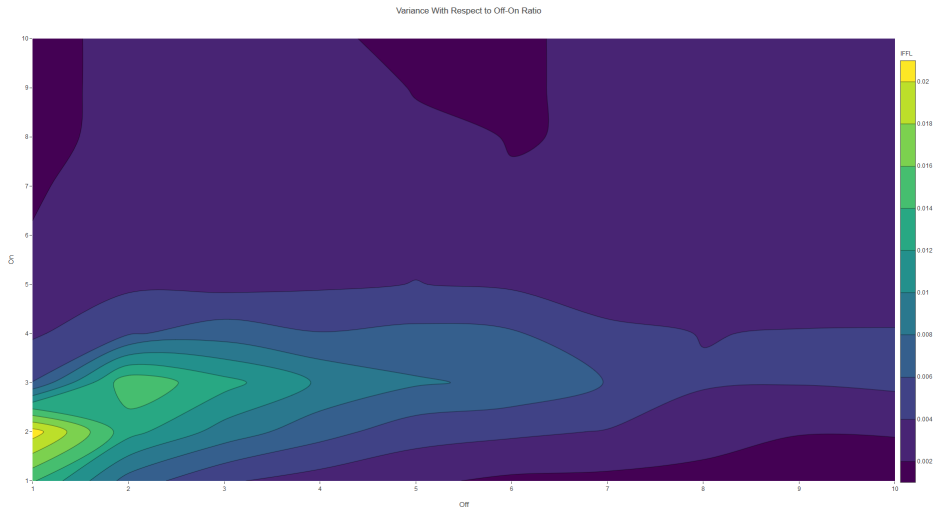


Figure. 6: Contour graph shows that at off-on ratio of 1:2, variability in concentration of expression is highest with respect to parameter value range. This variance is still much lower than that of naive model.

The demonstration of noise damping property of IFFL can be explained through proportionality effect of mf-Lon. Even with noisy inputs, mf-Lon is produced accordingly and weakens the effect of stochastic expressions seen in m-Scarlet. In the case of naive's high variance, result shows very consistent noise response through all combination of off and on ratios. Regardless, IFFL is exponentially more robust to noise than its naive counterpart.

3 Results

I. IFFL is better at distinguishing small changes in temporal structure, ie. small off-on ratio.
II. IFFL is better at processing time domain input when ratio of on to its period is large.
III. IFFL is exponentially more robust to noise than naive due to the proportionality effect of the protease on the reporter.

4 References

[1] Dolmetsch R.E., Xu K., and Lewis R.S., Calcium oscillations increase the efficiency and specificity of gene expression. Nature, 1998. 392(6679):

p. 933–6. pmid:9582075

[2] Batchelor E., et al., Stimulus-dependent dynamics of p53 in single cells. Mol Syst Biol, 2011. 7: p. 488. pmid:21556066

[3] Goentoro L., et al., The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol Cell, 2009. 36(5): p. 894–9. pmid:20005851

[4] Zhang C, Tsoi R, Wu F, You L (2016) Processing Oscillatory Signals by Incoherent Feedforward Loops. PLoS Comput Biol 12(9): e1005101. https://doi.org/10.1371/journal.pcbi.1005101

[5] http://2009.igem.org/Team:Aberdeen_Scotland/parameters/invest_1

5 Supplementary Materials

5.1 Model Code in Python

Portions of codes written for analysis can be found here: https://github.com/intj1/Math300/ Where Lorenz curves are used to illustrate the motif's output sensitivity relative to the naive circuit. This approach is not mentioned in this paper since the vector difference does a better job in elucidating IFFL's robustness.

5.2 Variance Test Table as Direct Translation to Noise Response

The data below can be replicated using the "Heat IFFL.ipynb" file: (https://github.com/intj1/Math300/blob/master/IFFL%20Heat/Heat%20IFFL.ipynb).

Off	On	IFFL Average Variance	Naive Average Variance
1	1	0.01533882340566167	0.13418271378415284
1	2	0.02187836097282671	0.13418271378415284
1	3	0.005754886844342153	0.13418271378415284
1	4	0.003958342560366722	0.13418271378415284
1	5	0.003155402960724867	0.13418271378415284
2	1	0.007574283020729188	0.13418271378415284
2	2	0.01389145087320129	0.13418271378415284
2	3	0.015895666169373285	0.13418271378415284
2	4	0.005935498143123905	0.13418271378415284
2	5	0.0037181405342439017	0.13418271378415284
3	1	0.004811724265443684	0.13418271378415284
3	2	0.00926761629811609	0.13418271378415284
3	3	0.013525597130485838	0.13418271378415284
3	4	0.007224533785821764	0.13418271378415284
3	5	0.0035019045270745872	0.13418271378415284
4	1	0.0031226758003070555	0.13418271378415284
4	2	0.007279214430683052	0.13418271378415284
4	3	0.010088694635714508	0.13418271378415284
4	4	0.006258872666223699	0.13418271378415284
4	5	0.00384901646694585	0.13418271378415284
5	1	0.0025600645664398554	0.13418271378415284
5	2	0.005072125710890843	0.13418271378415284
5	3	0.0087723345683548	0.13418271378415284
5	4	0.0067162680364338135	0.13418271378415284
5	5	0.004191180349174252	0.13418271378415284