Chapter 3

Correlated Noise in a Synthetic Gene Circuit Reveals Regulation

Simulations and analysis in the previous chapter are based on parameters for the λ cI system. Here, we test those predictions experimentally by building a synthetic circuit using the λ cI repressor and other characterized circuit elements in *E. coli* bacteria. The synthetic circuit uses three fluorescent proteins so that noisy gene expression can be monitored with time-lapse microscopy. To apply the analysis methods in the previous chapter, the cross correlation calculation is extended to allow for branched data where all progeny originate from a single ancestor. Experimental cross correlation curves show features that were predicted in simulation and results are tested further by comparing two different noise backgrounds to study the relative effect of extrinsic and intrinsic noise.

3.1 Synthetic Construct

We built a synthetic gene circuit with one transcription factor and three fluorescent proteins (Fig. 3.1). A protein fusion of the transcription factor λ cI and yellow fluorescent protein (YFP) represses production of red fluorescent protein (RFP), which is under the control of a variant of the λ P_R promoter, O_R2^* [41]. Cyan fluorescent protein (CFP) is controlled by a strong independent constitutive promoter, based on the σ_{70} consensus sequence. The construct was transformed into *E. coli* strain MG1655Z1, a derivative of MG1655 that overexpresses LacI [42]. The promoter for cI-YFP is controlled by LacI, which can be inactivated by the inducer IPTG. Thus, inducer concentration can be adjusted to place the mean cI-YFP concentration in an active regulatory range.

Two versions of this construct were built: (1) A chromosomally integrated version (single copy) and (2) a low-copy plasmid version ($\sim$10 copies). Fig. 3.1 shows the plasmid map; integration details are discussed in Section 3.6.2.

The design of this circuit is based upon work by R. Sidney Cox on a similar circuit, which is studied in Section 3.5. The fluorescent proteins are oriented to minimize read-through from one gene

to the next. In addition, multiple terminators are placed in between genes to stop transcriptional read-through. Each promoter in the circuit is bracketed by unique restriction sites to allow for easy interchange of promoters. A kanamycin antibiotic resistance marker is included to select for bacteria that contain the synthetic construct. In the chromosomally integrated version, the region from the kanamycin resistance marker through the end of cI-YFP was integrated into the *galK* site of MG1655Z1. The plasmid version uses a low-copy origin of replication, SC101 [42].

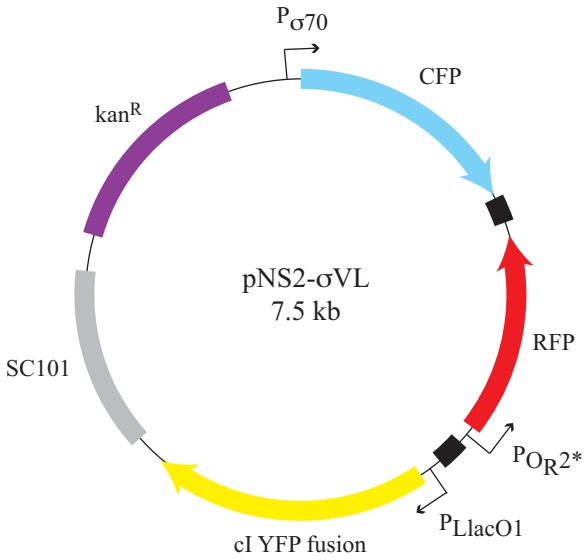

Figure 3.1: Plasmid version of the synthetic gene circuit. Colored arrows are genes, small black arrows are promoters. Black boxes indicate several terminators placed next to each other. The plasmid is a 7,500 base pair loop of DNA that replicates independently inside the cell.

A schematic representation of how noise affects the circuit is shown in Fig. 3.2. Each individual gene had intrinsic noise associated with it, while extrinsic noise affects all genes.

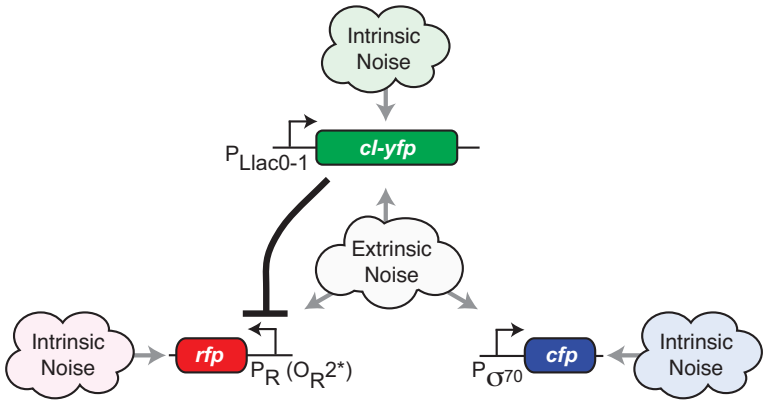

Figure 3.2: Noise sources affecting the three-color circuit.

3.2 Static Data

To quantify the behavior of the gene circuit in response to the inducer IPTG we took snapshots of individual cells and measured expression of the three fluorescent proteins (Fig. 3.3). As expected, CFP is relatively independent of IPTG, YFP increases as LacI is inhibited, and RFP decreases as cI-YFP increases. In the movies discussed below we used between 10–15 μM IPTG concentrations to ensure that cI-YFP was actively regulating RFP.

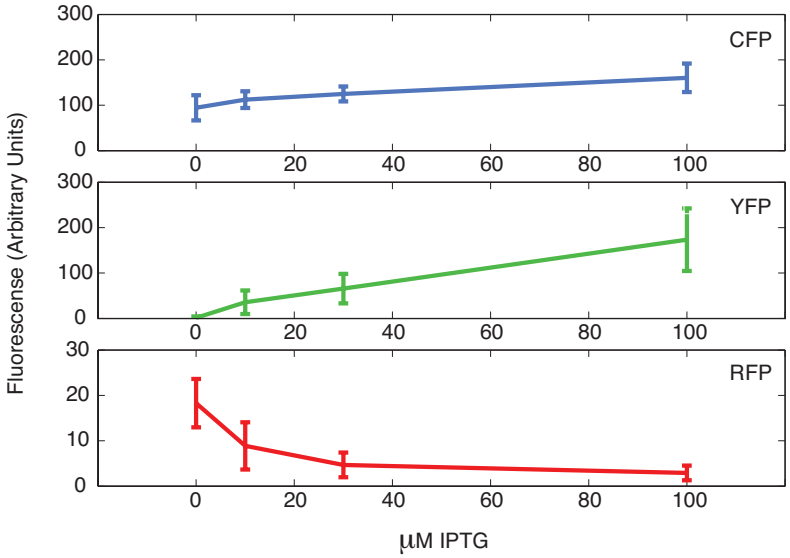

Figure 3.3: Dose response curves for three-color circuit with IPTG. Data comes from single cells (100–200 per data point), measured at 0, 10, 30, 100 μM IPTG. Error bars are the standard deviation of the data. Background has been subtracted from all fluorescence data.

We plotted snapshot data to show raw correlations between genes before considering temporal effects. Fig. 3.4 shows data from the plasmid and the chormosomally integrated versions of the construct; each dot is data from a single cell. Comparing the unregulated genes (CFP and YFP) with the regulated genes (RFP and YFP), a negative correlation can be seen due to repression of RFP by cI-YFP. The pair of unregulated genes are positively correlated, even in the absence of any regulatory elements linking them together. This correlation is the product of extrinsic noise in the system. Moving to the plasmid case, correlation between all pairs of genes increases. Furthermore, RFP and YFP on the plasmid show similar levels of correlation to CFP and YFP on the chromosome. It is clear that from these data alone it is not possible to infer the presence of regulation or details about its strength and the time scale on which it occurs.

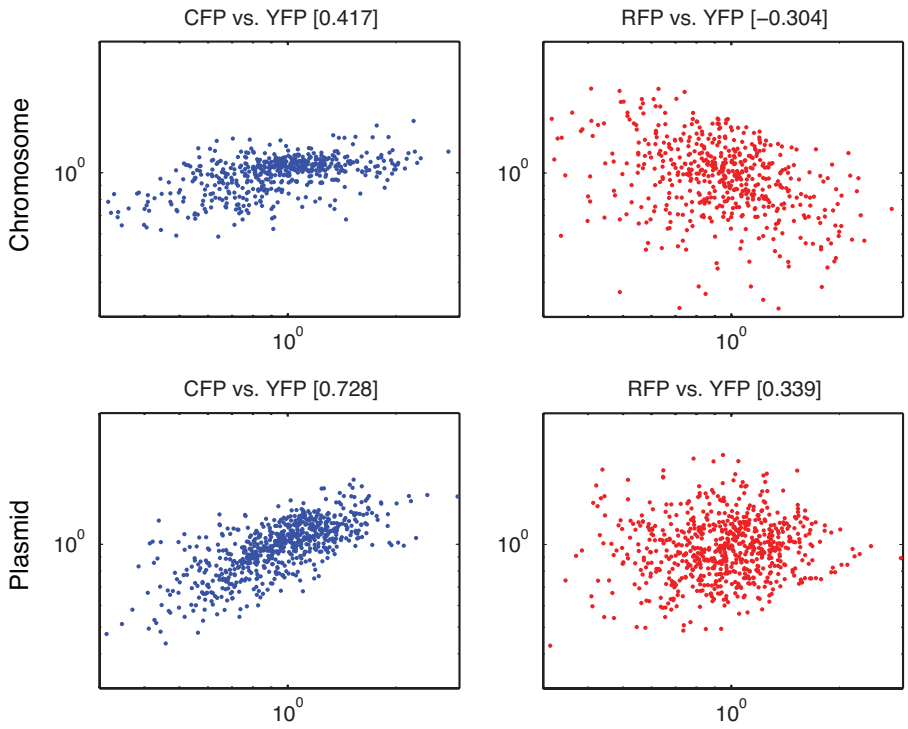

Figure 3.4: Expression of RFP, YFP, and CFP in individual cells plotted against each other. Title bars contain the correlation coefficient for these data. Plots are shown in log scale. Fluorescence data are background subtracted, units are arbitrary.

3.3 Temporal Data

Cells were grown and imaged in three colors using automated time-lapse fluorescence microscopy. A filmstrip from one movie of the chromosomally integrated circuit is shown in Fig. 3.5a. Although cells are imaged in all three colors at once, the two filmstrips show YFP & RFP and YFP & CFP separately for clarity. Note that pixels that contain equal levels of red and green appear as yellow, while equal levels of blue and green appear cyan. Individual colors are shown in grayscale in the right panel of Fig. 3.5a for a single time point. These images show strong anti-correlation between RFP (red) and YFP (green), while CFP (blue) is expressed at a more homogeneous level across all cells (see right panel of Fig. 3.5a). The expression of CFP does not appear to be linked to the expression of YFP or CFP, consistent with the design of circuit. The appearance of spatially grouped sub-populations of cells that display similar fluorescence states occurs because τ_{reg} exceeds the cell cycle time, consistent with simulation predictions (Fig. 2.5) and observations in [36].

A filmstrip of the alternate plasmid-based construct is shown in Fig. 3.5b. We expect an increase in extrinsic noise due to plasmid copy number fluctuations. Thus, the correlations between all three colors should be higher than in the chromosomal case. Although anti-correlation between RFP (red) and YFP (green) is visible, there is a marked increase in the number of yellow pixels, indicating

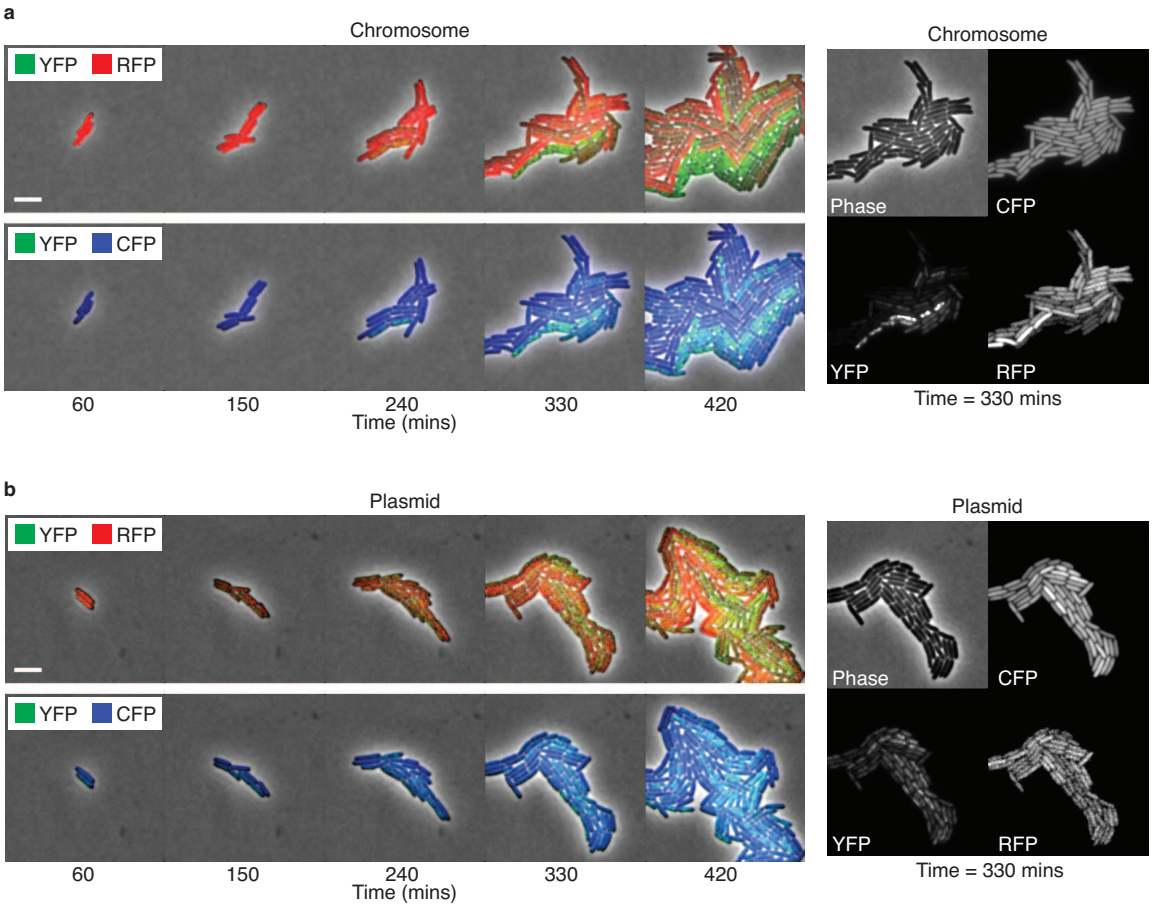

Figure 3.5: Time-lapse movies of gene expression fluctuations in a synthetic genetic circuit. (a) Left panels show filmstrip of cells with chromosomally integrated circuit. Note strong anti-correlation between RFP (red) and YFP (green), and the lower correlation between CFP (blue) and YFP. Scale bar, 5 μm. Right panel shows individual colors and phase images for $t = 330$ mins. Note anti-correlation between RFP and YFP and the uniform expression of CFP. (b) Filmstrip of the same circuit on a low copy plasmid shows increased variability in all colors. In particular, in the right panel note increased variability of CFP relative to the chromosomal case. Colors and scale bar are the same as in (a). Extended movies are shown in Appendix B.

increased correlation between these two genes. Comparing the black and white panels for CFP in Fig. 3.5a and b, the plasmid shows increased variability in this constitutive promoter over the chromosomal case. The right panel also shows the effects of extrinsic noise, where cells that are bright in one color tend to be bright in all three colors.

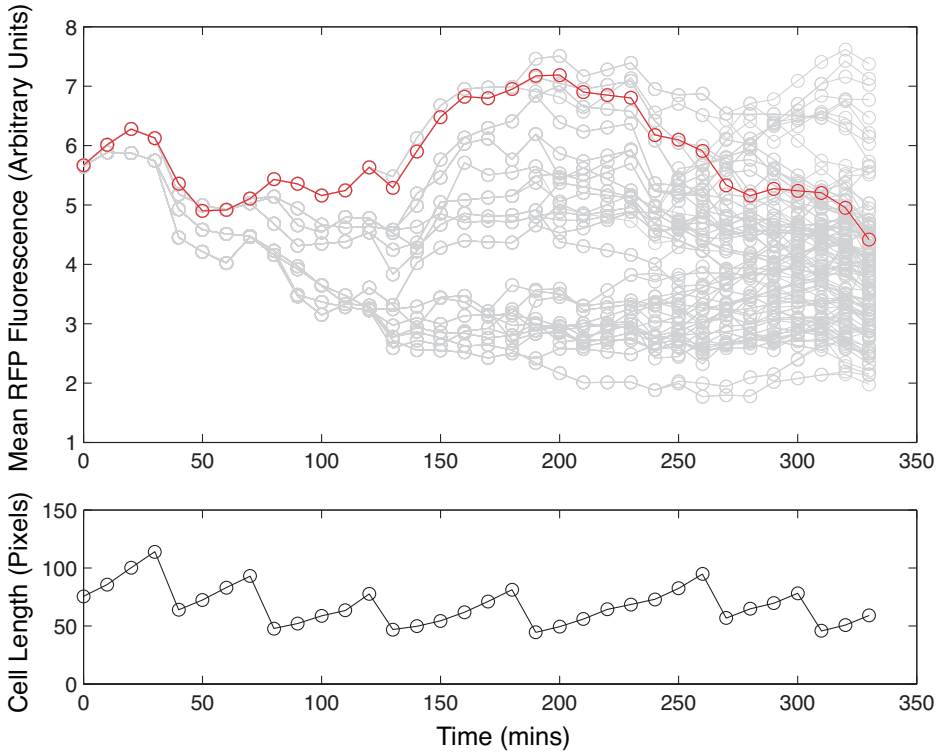

Figure 3.6: Mean (averaged over the cell) expression of RFP for all cells in a growing microcolony. A single representative trace is highlighted in red. Data on cell length for this representative cell lineage are shown in the bottom panel. Decreases in the cell length are cell division events. Note that these decreases are not visible in the RFP data because it has been averaged by cell volume.

To analyze these data quantitatively, we used semi-automated image analysis software to extract fluorescence intensities for individual cells across the lineage tree of a growing microcolony (Fig. 3.6). These data show a wide diversity of expression levels at the end of the movie. A single trace is highlighted in red and the corresponding cell length is shown in the bottom panel—variations in gene expression are not strongly linked to the cell cycle. Typical time traces from a single cell lineage for all three fluorescent genes are shown in Fig. 3.7. More variability is seen in the expression of plasmid-based genes than in their chromosome-based counterparts, consistent with the enhanced level of extrinsic noise. The same data shifted in time reveal temporal anti-correlation between cI-YFP and RFP signals (Fig. 3.8).

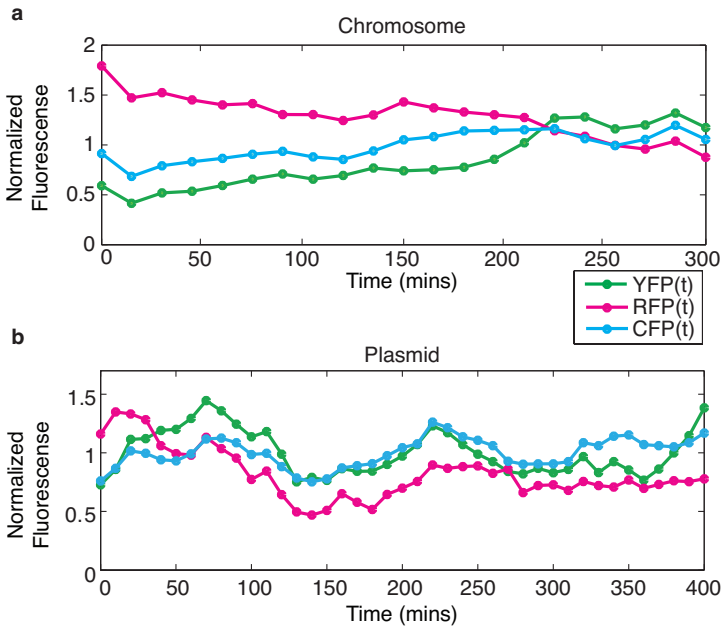

Figure 3.7: Typical lineage traces show noise in gene expression. Data are normalized by mean intensity.

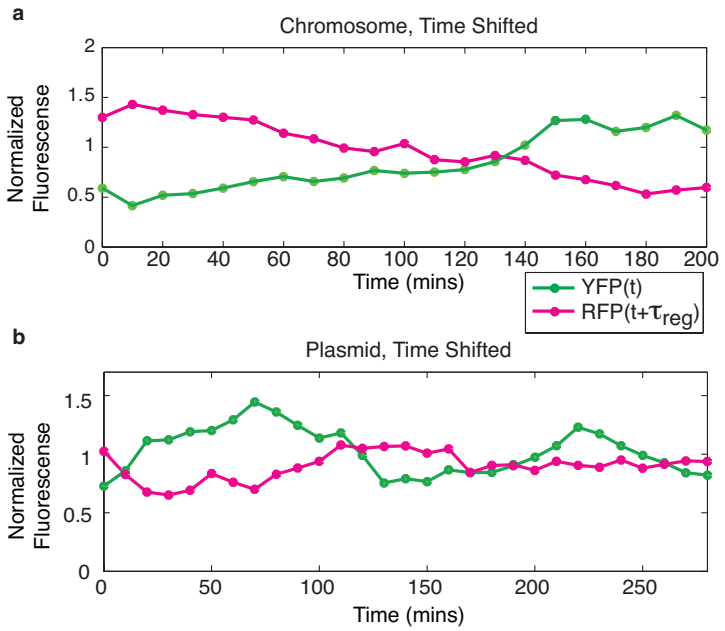

Figure 3.8: Data are the same as in Fig. 3.7, however RFP data are shifted to reveal the delayed anti-correlation. (chromosome: $\tau_{reg} = 100$ mins, plasmid: $\tau_{reg} = 120$ mins)

3.4 Experimental Cross Correlations

3.4.1 Extension of the Cross Correlation Function to Branched Data

The cross correlation function was calculated using these temporal data. However, it was necessary to introduce a modified formula for the cross correlation function to properly account for the branching nature of the data. Fig. 3.9 illustrates the need for an extension to the traditional cross correlation formula. In this example we start with one cell, which divides into two cells, and finally four cells. Although we have four time-series measurements, some of the data between them are shared. Thus, data from the trunk of the tree are counted four times, while data at the leaves are only counted once. This is potentially problematic if the characteristics of the correlation at early times are different from those at the end; the correlation expression will be heavily biased towards the early data. To correct for the problem of over-counting data we introduced a modified expression for the

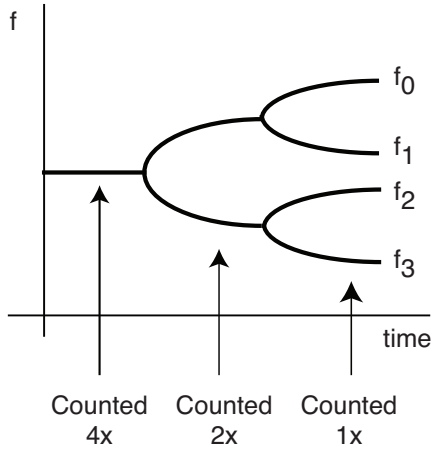

Figure 3.9: Cross correlations on data with a tree structure need to avoid over-counting.

cross correlation function, $R_{f,g}(\tau)$, given by

$$
\begin{aligned}
S_{f,g}(\tau) &= \begin{cases} \frac{1}{N-|\tau|}\frac{1}{N_{cells}} \left[\sum_{i=0}^{N_{cells}-1}\left(\sum_{n=0}^{N-\tau-1}\tilde{f}_i(n+\tau)\tilde{g}_i(n)\right) - \sum_{i=0}^{N_{cells}-2}\left(\sum_{n=0}^{k_i-\tau-1}\tilde{f}_i(n+\tau)\tilde{g}_i(n)\right)\right] & \tau \geq 0 \\ S_{g,f}(-\tau) & \tau < 0 \end{cases} \\
R_{f,g} &= \frac{S_{f,g}(\tau)}{\sqrt{S_{f,f}(0)S_{g,g}(0)}}.
\end{aligned} \tag{3.1}
$$

This expression has two terms: The first term is the standard expression for the cross correlation function, averaged over the number of data traces (N_{cells}). The second term accounts for any over-counting of data by subtracting data that have been counted more than once.

N_{cells} is the total number of cells at the final time point (four in Fig. 3.9) and k_i is the branching

point between traces f_i and f_{i+1}, described in further detail below. Other terms are the same as defined in the original cross correlation expression given in Eqn. (2.16) except the mean subtraction is extended to the branched situation using

$$\tilde{f}_i \quad = \quad f_i - \sum_{i=0}^{N_{cells}-1} f_i.$$

Fig. 3.10 further illustrates the data analysis process. The formula in Eqn. (3.1) assumes that the numbering of the data traces is ordered in a manner so that, for all data, each pair of adjacent data traces f_i and f_{i+1} share as much data as possible. Fig. 3.10a illustrates numbering before correction. A preprocessing step must be applied to renumber these data traces, as in Fig. 3.10b. The branching points between data are listed on the x-axis.

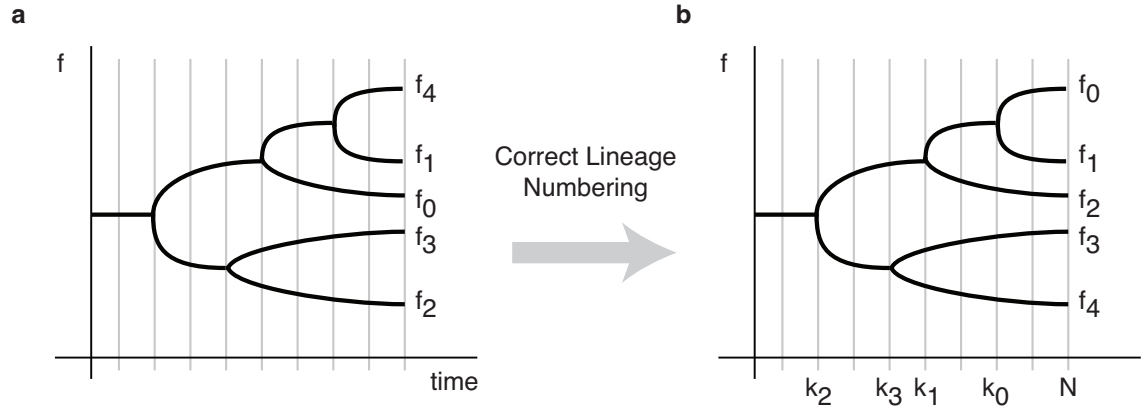

Figure 3.10: Tree data must be preprocessed for optimal numbering. Here $N_{cells} = 5$, $N = 9$, $k_0 = 7$, $k_1 = 5$, $k_2 = 2$, $k_3 = 4$.

3.4.2 Experimental Noise Correlations

Fig. 3.11 shows the resulting cross correlations for cases of active repression (cI-YFP and RFP) and no regulation (cI-YFP and CFP). These functions displayed all features predicted by the model, including a strong dip at a negative lag time due to repression, and positive correlation at zero lag due to global noise in the unregulated case.

Comparing the chromosomal and plasmid-based constructs shows how the relative amplitude of intrinsic and extrinsic noise affects cross correlation functions. Recall that copy number fluctuations in the plasmid increase the effective extrinsic noise level for the genes in the circuit, and reduce the relative importance of intrinsic noise, whose uncorrelated fluctuations average out. Comparing Figs. 3.11a and b shows that the amplitude at $\tau = 0$ is increased on the plasmid relative to the chromosomal construct in both the regulated and unregulated case, reflecting simultaneous correlations. These results confirm model predictions, and demonstrate that regulation can be discriminated even

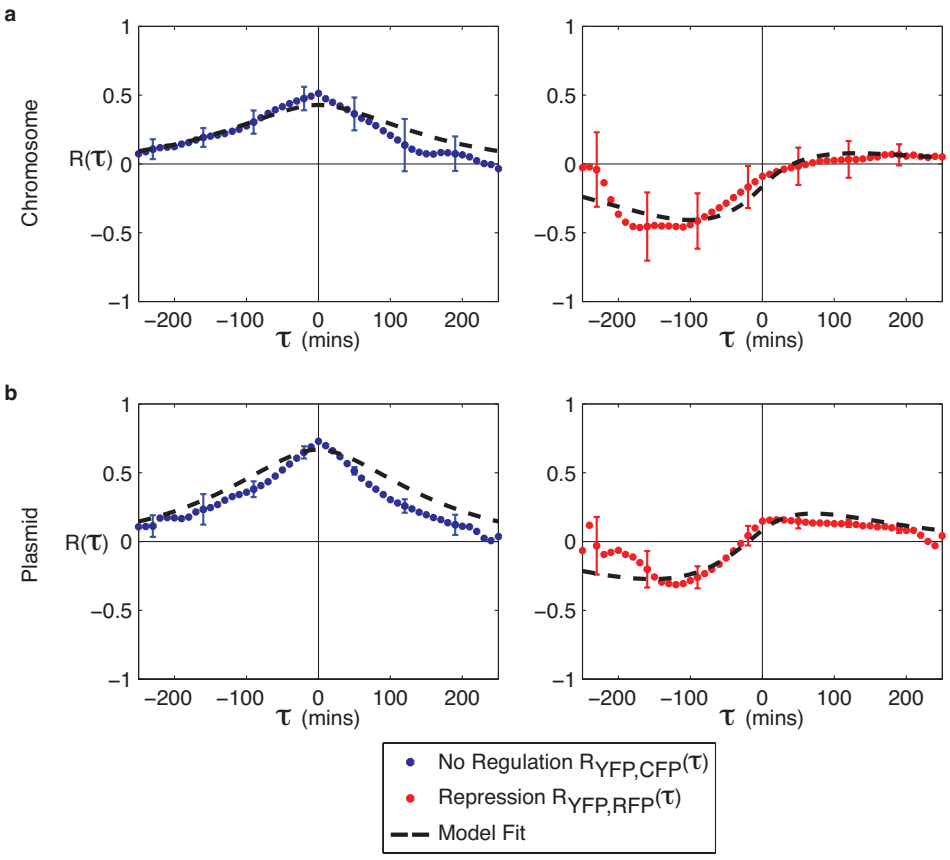

Figure 3.11: Experimental cross correlation functions for chromosomal and plasmid-based circuits. Chromosomal data are averaged over $n = 5$ independent movies (with 100–200 cells per microcolony upon movie completion), $n = 6$ movies for plasmid data. Error bars show standard error of the mean. (a) Cross correlation functions $R_{YFP,CFP}(\tau)$ (blue circles) and $R_{YFP,RFP}(\tau)$ (red circles) and model fits (dashed black line) for the chromosomally integrated construct. For the model fit to the chromosomal construct, $g = -0.01$, $W_E/W_I = 4.5$, and $\alpha_B = 1.7$. (b) Cross correlation functions for the plasmid-based construct. For the model fits, $g = -0.01$, $W_E/W_I = 1.7$, $\alpha_B = 0.5$.

when extrinsic noise amplitudes are large.

The model fits shown in Fig. 3.11 were generated using MATLAB optimization software to minimize the difference between experimental and analytical cross correlation expression from Eqns. (2.26)–(2.26), taking into account standard errors for experimental data points. Three parameters were fit: g, the derivative of the gene regulation function evaluated at the steady-state repressor concentration (see Chapter 2); W_E/W_I, the ratio of extrinsic to intrinsic noise; and α_B, the rate of protein production at steady-state. All other parameters are listed in Table 2.1.

The autocorrelation curves, $R_{x,x}(\tau)$, can also be calculated using these experimental measurements. By using the analytic methods described in Chapter 2 we can calculate expressions for the autocorrelation for the two extreme cases where only extrinsic or intrinsic noise is present. Since our system is influenced by both noise sources, we expect the actual data to fall within these theoretical bounds. With only extrinsic noise acting on a signal $x(t)$ we model gene expression by

$$\begin{aligned} \dot{x} &= -\beta x + E \\ \dot{E} &= -\beta E + \eta_e. \end{aligned}$$

Using the analytic methods described in Chapter 2 we find the normalized autocorrelation expression

$$R^{ext}_{x,x}(\tau) = \exp^{-\beta|\tau|}\left(1 + \beta|\tau|\right).$$

With only intrinsic noise the model is

$$\begin{aligned} \dot{x} &= -\beta x + I_x \\ \dot{I}_x &= -\kappa I_x + \eta_x, \end{aligned}$$

finding the normalized autocorrelation to be

$$R^{int}_{x,x}(\tau) = \frac{\kappa \exp^{-\beta|\tau|} - \beta \exp^{-\kappa|\tau|}}{\kappa - \beta}.$$

Note that if $\kappa \gg \beta$ this expression simplifies to $\exp^{-\beta|\tau|}$, as in [36].

The autocorrelation curves for the plasmid data (Fig. 3.12) are bounded by the theoretical limits for autocorrelations due to solely extrinsic or intrinsic noise.

3.5 Alternative Construct

These cross correlation calculations were applied to data from a similar three-color circuit that was built by R. Sidney Cox. The goal of building this circuit, shown in Fig. 3.13, was to create

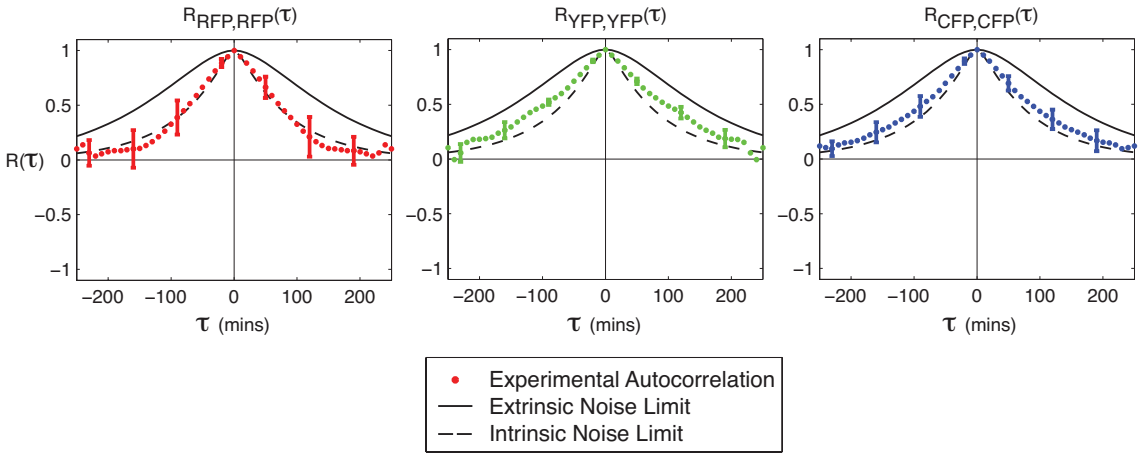

Figure 3.12: Autocorrelation curves for the plasmid data and theoretical noise limits. The parameters $\beta = \log(2)/T_{cc}$ and $\kappa = \log(2)/T_{int}$ are listed in Table 2.1. Error bars are standard error.

a construct for easy measurement of gene expression from three promoters, $P1$, $P2$, and $P3$. To test properties of the circuit, several well-characterized promoters, TetR, LacI, and a combinatorial promoter with LacI and AraC, were placed in the construct. Cox was able to measure cross-talk between the promoters, read-through between genes, and measure noise under different levels of induction for each promoter (R. S. Cox, personal communication).

Although certain features are discernible from standard correlations between genes, temporal correlations contain more information. Fig. 3.14 shows (1) higher correlation between YFP and RFP than the other pairs of genes due to the co-regulation by LacI, (2) correlations between YFP and CFP and CFP and RFP are similar in magnitude and are the result of extrinsic noise in the system, (3) dynamics governing CFP expression and regulation are slower, causing extended periods of positive correlation, (4) all three curves are centered at zero, indicating that no uncharacterized regulation is present.

3.6 Methods and Characterization

3.6.1 Bleaching Times

We measured bleaching times for each of the three proteins in the new synthetic construct. Fig. 3.15 shows the response of fluorescent proteins to prolonged exposure to measurement conditions. These data were gathered by preparing agarose pads with dense cells on them and exposing them to fluorescent light for 10 minutes. After each minute, a snapshot was taken of the cells in all three colors. Exposure times of these measurement snapshots were short—typically less than a second—and are not expected to significantly affect the measurements.

Cells in each column of Fig. 3.15 were bleached in a single color. Rows are measurements of the

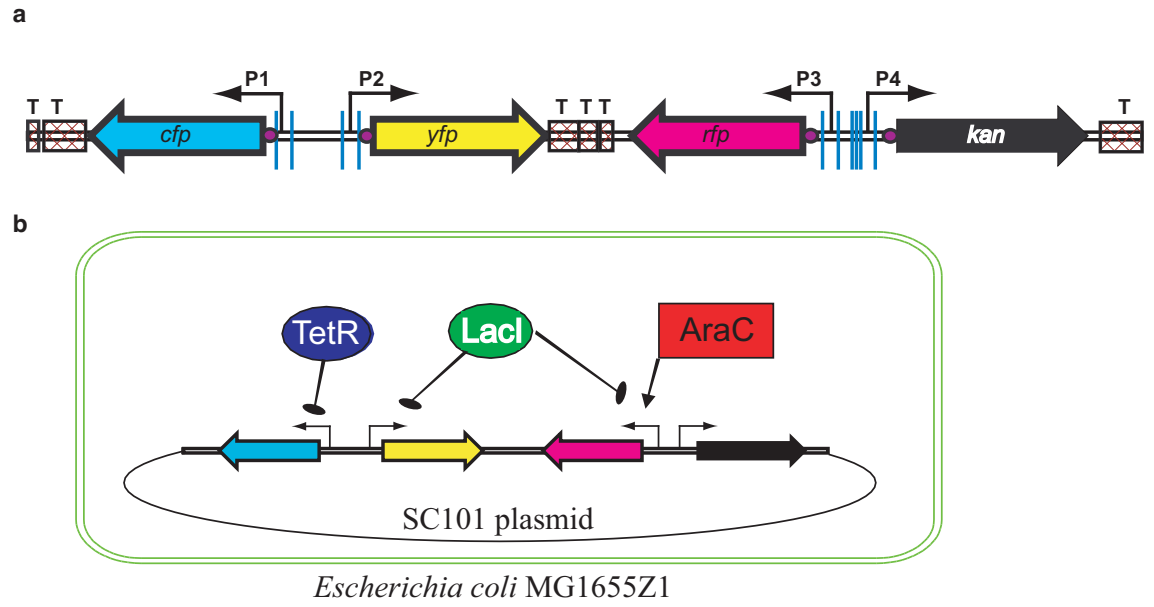

Figure 3.13: Alternate three-color circuit. (a) Construct for measuring expression from three promoters. Terminators are labeled "T," purple dots are ribosome binding sites, blue lines indicate unique restriction sites for cloning. (b) Promoters used in data analysis. The promoter on RFP is a combinatorial promoter that is repressed by LacI and activated by AraC. Figure courtesy of R. Sidney Cox.

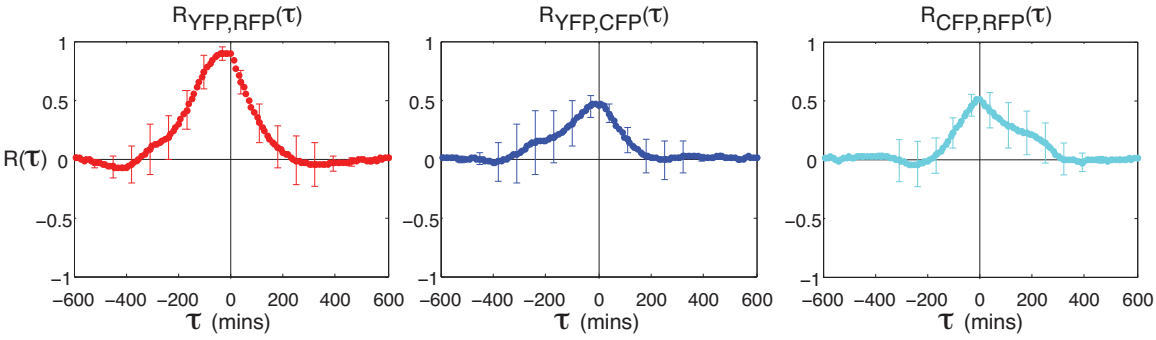

Figure 3.14: Cross correlations between all pairs of genes in the synthetic circuit shown in Fig. 3.13. Error bars are standard error.

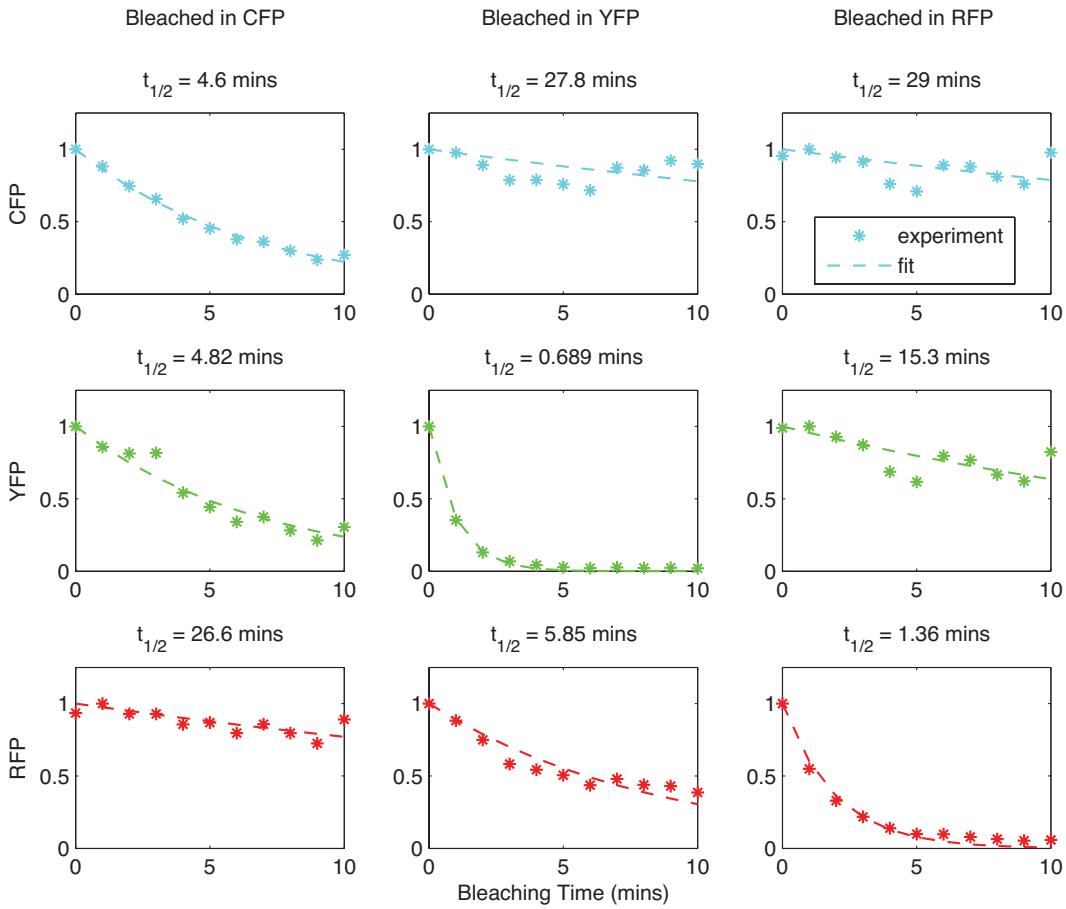

Figure 3.15: Bleaching times for CFP, YFP, and RFP. Times to half signal are shown in the title bars.

three fluorescent protein levels in response to the bleaching. The plots along the diagonal show how the measured signal degrades with time. YFP is the biggest concern for bleaching, with the signal halving every 40 seconds. CFP is remarkably resistant to bleaching. Off-diagonal plots indicate how fluorescence in one channel degrades due to exposure in a different channel. Again, YFP is sensitive to bleaching even for exposure in other channels. For each set of data, a decaying exponential was fit to the curve and the half-life of the protein is listed in the title of each subplot.

Exposure times in movie conditions are typically on the order of 1 second or less per color, taken once every 10 minutes. At these levels bleaching is not expected to be a significant concern.

3.6.2 Methods

The plasmid pNS2-σVL (Fig. 3.1) was constructed by synthesizing a region starting with the kanamycin promoter and ending just before cI-YFP (synthesis by Blue Heron). The sequences for CFP and the red fluorescent protein mCherry were modified as in GFPuv116 of [43] in regions immediately downstream of the start codon, and were codon-optimized for expression in *E. coli*. The synthesized construct was cloned into the plasmid pZS21-cIYFP [19], replacing the Tet promoter with the synthesized fragment.

The strong promoter on CFP was based on the σ_{70} consensus sequence:

$$5' - \texttt{aataattcTTGACAtttatgcttccggctcgTATAATacgtgcaatt} - 3',$$

where the capital letters highlight the location of the -35 and -10 boxes.

To integrate the construct into the chromosome, the region from the kanamycin resistance marker through the end of cI-YFP was amplified using PCR with 50 base pair homology regions for galK

$$\text{H1:} \quad 5' - \texttt{TTCATATTGTTCAGCGACAGCTTGCTGTACGGCAGGCACCAGCTCTTCCG} - 3'$$

$$\text{H2:} \quad 5' - \texttt{GTTTGCGCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGTAAGAA} - 3'$$

included on the end of each primer. The construct was integrated into the *galK* region of MG1655Z1, using the recombineering methods described in [44]. Insertion was verified with colony PCR.

Single colonies were inoculated in selective LB media and grown overnight. This culture was diluted back 1:100 in 1/4 strength LB with 30 μg/ml kanamycin and between 10–15 μM IPTG (varies for different movies). The cells were then grown to OD 0.1–0.2 and diluted back 1:100 in M9 minimal media containing 0.2% glycerol, 0.01% Casamino Acids, 0.15 μg/ml biotin, and 1.5 μM thiamine (we denote this media MGC). Cells were placed on 1.5% MGC low melting temperature agarose pads containing 10–15 μM IPTG and grown at 37°C for 3 hours to equilibrate to the inducer conditions on the pad. The pad was then placed in 200 μl of MGC + IPTG and shaken to release the

cells. These equilibrated cells were placed on a fresh pad for time-lapse imaging. The temperature of the microscope chamber was kept at 32°C for the duration of the movie. Images were taken every 10 minutes in phase and each of the three fluorescent color channels. Fluorescence analysis of cell lineages with done with custom MATLAB software.