SYNCHRONIZATION OF STOCHASTIC CIRCADIAN CLOCK OSCILLATORS

by

LINGYUN WU

(Under the Direction of Heinz-Bernd Schüttler)

ABSTRACT

Circadian clock oscillators in living organisms have been studied extensively for decades. However, it is presently still not fully understood how the stochastic clock gene regulatory processes in individual cells are getting organized to produce coherent, nearly periodic clock oscillations observed at the level of large cell populations. To investigate the effects of intracellular stochasticity, arising from the stochastic gene regulation at the multi-cell level, we have developed a hybrid model, which combines the random clock gene flip processes with a deterministic time evolution of resulting gene products and signaling agents, subject to a quorum sensing-type inter-cellular coupling mechanism, for the microbial fungus *Neurospora crassa*. The cell population, with sufficient coupling strength, can generate coherent and nearly periodic average signals, whereas the dynamics of each single cells still remains irregular and incoherent. We introduce several novel non-linear time series analysis methods to study the underlying mechanism of this paradoxical collective system behavior. By using the relative signal amplitude as the synchronization order parameter, a continuous phase transition phenomenon can be observed in the system, from being non-oscillatory to oscillatory, with the population size approaching infinity.

INDEX WORDS: Monte Carlo simulations, biological clock, stochastic gene expression, Neurospora crassa, quorum sensing, synchronization, phase transition

SYNCHRONIZATION OF STOCHASTIC CIRCADIAN CLOCK OSCILLATORS

by

LINGYUN WU

B.S., Shandong University, China, 2012

B.S., Portland State University, 2012

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

Synchronization of Stochastic Circadian Clock Oscillators

by

Lingyun Wu

Major Professor:	Heinz-Bernd Schüttler
Committee:	David P. Landau Steven P. Lewis

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
December 2020

This dissertation is dedicated to

my family and my wife, Jiaxin

for their constant support and unconditional love.

Acknowledgments

First of all, I would like to sincerely acknowledge my advisor, Dr. Heinz-Bernd Schüttler, for his invaluable instructions, generous support, immense knowledge, and great patience throughout my Ph.D. studies. His guidance helped me in all the time of research and writing of this dissertation.

In addition, I would like to express my appreciations to Dr. Jonathan Arnold for his encouragement, support, and unlimited help to my research. I'm also grateful for my committee members, Dr. David P. Landau and Dr. Steven P. Lewis, for providing me with very helpful suggestions and comments to my dissertation.

Furthermore, I would like to thank Dr. Cristian Caranica, Dr. Zhaojie Deng, Jia Hwei Cheong, and Xiao Qiu for the excellent collaboration in my research. I also truly appreciate many members of the CSP. I benefited a lot from the discussion with Dr. Guangjie Shi, Dr. Kai Qi, Dr. Jiahao Xu, Dr. Ziwei Zhang, Kellan Gibson and Shengming Zhang. I also wish to thank Dr. Shan-Ho Tsai, Mike Caplinger and Jeff Deroshia for being always so patient and willing to help me with all kinds of problems with computing facilities.

Finally, I would like to acknowledge the support of NSF MCB-1713746 from NSF Systems and Synthetic Biology (SSB) and NSF Physics of Living Systems (PoLS).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Circadian clocks are biological oscillators that regulate many physiological behaviors of living systems from cellular level to organismal level [1]. It can generate well self-sustained oscillations with a period of around 24 h without outside drivers [2,3] or be entrained by environmental stimuli, such as light and temperature [4–8]. On the single-cell level, the system dynamics of a circadian system is generally developed in the form of genetic networks, which provides an explicit framework of gene, RNAs, proteins, and, more importantly, reactions that connect them [9]. However, stochastic fluctuations, due to the low number of biological molecules, are quite unavoidable in modeling clock oscillators. Gene expression noise, which is mostly caused by random changes of genes' transcriptional states and their low copy numbers, becomes a significant part of noise source in the single-cell oscillator system [10].

My research mainly focuses on the circadian system in the filamentous fungus, *Neurospora crassa*, which has been well studied, both experimentally and theoretically, for decades. Figure 1.1 [11] shows a detailed genetic network of a single cell's biological clock module in *Neurospora crassa*. The system contains 26 reactions and 16 participating biomolecular species, which involves genes of *white-collar-1 (wc-1)*, *white-collar-2 (wc-2)*, *frequency (frq)*, and *clock controlled gene (ccg)* and their associated products [12]. The central feature of the genetic network that permits oscillations is a closed-feedback loop of biomolecular reactions which consists of protein complex *White Collar Complex (WCC)* formed by proteins *WC-1* and *WC-2*, *frq*-gene, which is transcriptionally regulated by *WCC*, and protein *FRQ*, which catalyzes the destructive reaction of *WCC*. In addition, a positive feedback loop which involves the activation of *wc-1* mRNA at reaction C1 is predicted not to play a essential role in the occurrence of oscillations, but it might help keep oscillations sustained by increasing the lifetime of *wc-1* mRNA with an as-yet-unexplained mechanism [11].

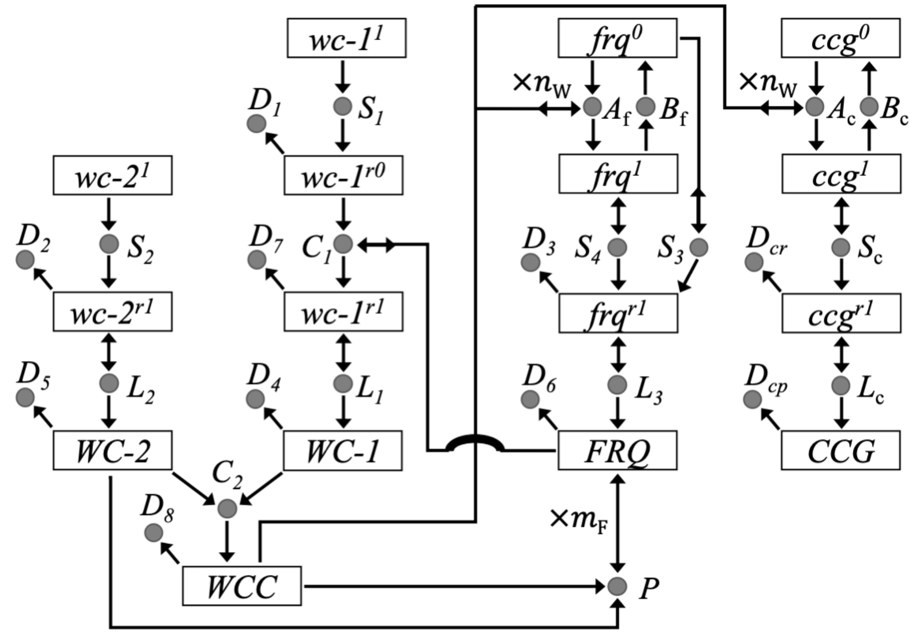

Figure 1.1: Genetic network for a single cell's the biological clock of *Neurospora crassa.* Boxes represent molecular species. Species which have superscripts 0 (1) and r0 (r1) indicate, respectively, transcriptionally inactive (active) genes and translationally inactive (active) mRNAs. Species with uppercase letters denote proteins. All biomolecular reactions are represented by circles, labeled by their reaction rates. Reactants are identified by arrows entering circles, products are identified by arrows leaving circles, and catalysts are identified by bidirectional arrows. In addition, circles without leaving arrows represent decay reactions, and $\mathrm{n_W}$ and $\mathrm{m_F}$ are Hill coefficients for reactions $\mathrm{A_f}$ and P.

A deterministic kinetic rate equation system, which is well constructed on the basis of the network shown above, has been successful in describing the biological clock dynamics at the level of large cell populations [11]. The concentrations of all molecular species, including *frq*-gene and *ccg*-gene, in the deterministic model are taken as continuous variables, since the large cell population size can considerably reduce the influences of stochastic fluctuations. However, it is presently an open question how the stochastic gene regulation and expression dynamics at the single-cell level can give rise to the coherent, periodic clock oscillations observed at the multi-cellular level.

In Chapter 2, I will introduce a hybrid model which uses a generalization of Gillespie kinetic Monte Carlo process to embed the stochastic gene flip events into the deterministic kinetic rate equation system based on a simplified reaction network of a single-cell clock in *Neurospora crassa*. A light-entrainment version of the hybrid model is described in this section, as well. Chapter 3 contains a multi-cell hybrid model which uses a quorum sensing-type signaling mechanism to couple cells in a population. In Chapter 4, several novel non-linear time series analysis will be introduced to study the basic physics of the special collective synchronization mechanism in the multi-cell hybrid model and its parametric robustness of the oscillatory behavior of the system. Chapter 5 presents a brief summary and conclusion.

Chapter 2

Hybrid model for the biological clock module in single cells of Neurospora crassa

2.1 The hybrid model with a simplified reaction network

In this section, a hybrid model which combines stochastic gene regulation of *frq*-gene and *ccg*-gen with deterministic time-evolution of the RNA and protein gene products is explicitly described. Figure 2.1 shows a simplified single-cell system in the dark, which is modified on the basis of the full genetic network for the deterministic model, shown in Figure 1.1. The protein *WC-2* in the original network is not considered as a key feature for oscillations of the clock, and, therefore, species, $wc\text{-}2^1$, $wc\text{-}2^{r1}$, and *WC-2*, are removed from the network. Furthermore, $wc\text{-}1^1$ is also taken from the network, for the reason that the number of gene *wc-1* keeps constant in the model. Consequently, the total number of species in this simplified network is reduced to 12, and we used it as the framework for the hybrid model.

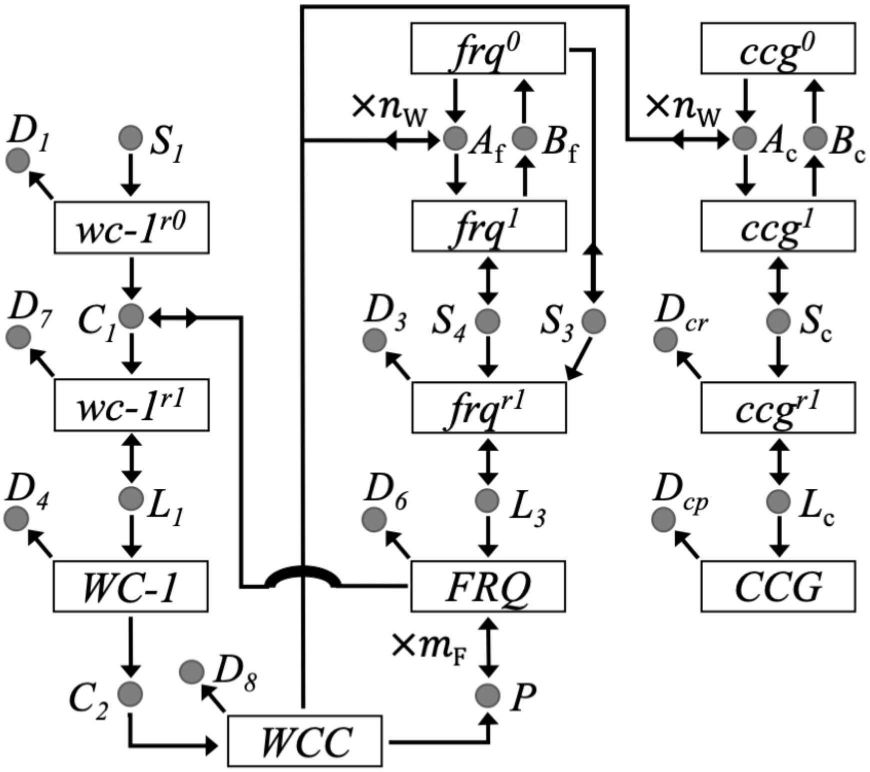

Figure 2.1: A simplified reaction network for the biological clock of Neurospora crassa

In the deterministic model, gene *frq* and *ccg* are described as continuous concentration variables for both two transcriptional states. However, assuming that there is only one copy of gene *frq* and *ccg* for each cell, their activation states should be binary variables with appropriate non-dimensionalization. The corresponding values for them are 1 and 0 for being transcriptionally active and inactive, respectively. Therefore, instead of using continuous variables for the four species, frq^0, frq^1, ccg^0, and ccg^1, we would like to replace them with two binary variables, denoted, respectively, by $g_{f,k}$ and $g_{c,k}$, where f and c in the subscripts refer to gene *frq* and *ccg* and k in the subscripts means that the k-th gene flip has occurred in the cell at time $t_k^{(gf)}$.

In the hybrid model, the process of flipping gene *frq* or *ccg* states is fully controlled by random WCC-binding and -unbinding events at the two genes' activator binding sites. Each time when a gene is randomly picked to be flipped in the cell, n_{W} WCC protein molecules are bound to or released from it. Obviously, these random gene flip events do cause discontinuities in the time evolution of the system's kinetic rate equations. The following procedures show how a generalization of the Gillespie kinetic Monte Carlo (KMC) process [13, 14] is employed to incorporate this discrete, stochastic gene flipping dynamics into the hybrid model, as follows:

(i) The time step, τ, between two successive gene flip events at $t_k^{(gf)}$ and $t_{k+1}^{(gf)} = t_k^{(gf)} + \tau$ is randomly chosen by using a uniformly distributed random variable, u, ranging in (0,1] in the Gillespie criterion,

$$\Psi(t = t_{k+1}^{(\mathrm{gf})} \mid t_0 = t_k^{(\mathrm{gf})}) \equiv \int_{t_k^{(\mathrm{gf})}}^{t_k^{(\mathrm{gf})}+\tau} dt' \sum_{\mathrm{x}\in\{\mathrm{f},\mathrm{c}\}} \Gamma_{\mathrm{x}}(g_{\mathrm{x},k}, t') = -\ln(u) \ , \tag{2.1}$$

where $\Psi(t \mid t_0 = t_k^{(\mathrm{gf})})$ is the integral of $\Gamma_{\mathrm{x}}(g_{\mathrm{x,k}}, t')$ from $t_0 = t_k^{(\mathrm{gf})}$ to t, and $\Gamma_{\mathrm{x}}(g_{\mathrm{x},k}, t')$ is the effective rate coefficient for flipping the transcriptional state of gene $\mathrm{x} \in \{\mathrm{f}, \mathrm{c}\}$ from $g_{\mathrm{x},k}$ to $g_{\mathrm{x},k+1} = 1 - g_{\mathrm{x},k}$ at time t'. The equation of $\Gamma_{\mathrm{x}}(g_{\mathrm{x},k}, t)$ is shown below,

$$\Gamma_{\mathrm{x}}(g_{\mathrm{x},k}, t) \ = \ \delta_{g_{\mathrm{x},k},0} \ A_{\mathrm{x}}([WCC]_t)^{n_{\mathrm{W}}} \ + \ \delta_{g_{\mathrm{x},k},1} \ B_{\mathrm{x}} \ , \tag{2.2}$$

where $\delta_{g,0}$ and $\delta_{g,1}$ are Dirac delta functions, $[WCC]_{t'}$ is the concentration of protein WCC at time t', and A_{x} and B_{x} are reaction rate coefficient with $\mathrm{x} \in \{\mathrm{f}, \mathrm{c}\}$.

(ii) To draw the random τ-value, for taking the time step from $t_k^{(gf)}$ to $t_k^{(gf)}$ in our algorithm, we therefore have to draw a uniform random number u from (0,1] and then solve Eq. 2.1 for τ. To do so, we need to evaluate the time integral, $\Psi(t, t_k^{(gf)})$, on the left-hand side (LHS) of Eq. 2.1. We therefore need to know the flipping rate coeffs,

$\Gamma_x(g_{x,k}, t)$, as functions of time, for times $t \geq t_k^{(gf)}$. Because of Eq. 2.2, this requires us to use a numerical ordinary differential equation (ODE) solver to solve the deterministic rate equations, starting at $t = t_k^{(gf)}$, for all the continuous non-gene species concentrations, including $[WCC]_t$. This ODE solution proceeds in time until the solution for τ in Eq. 2.1 is found. For numerical convenience, we include $\Psi(t, t_k^{(gf)})$ as one of the dynamical variables in the ODE system. We then use the ODE solver to also evaluate $\Psi(t, t_k^{(gf)})$, together with all the continuous non-gene species concentrations. The complete rate equation system to be solved, starting at $t = t_k^{(gf)}$, is given as follows:

$$\frac{d[wc\text{-}1^{r0}]_t}{dt} = S_1 - D_1 \cdot [wc\text{-}1^{r0}]_t - C_1 \cdot [wc\text{-}1^{r0}]_t \cdot [FRQ]_t \tag{2.3}$$

$$\frac{d[wc\text{-}1^{r1}]_t}{dt} = C_1 \cdot [wc\text{-}1^{r0}]_t \cdot [FRQ]_t - D_7 \cdot [wc\text{-}1^{r1}]_t \tag{2.4}$$

$$\frac{d[WC\text{-}1]_t}{dt} = L_1 \cdot [wc\text{-}1^{r1}]_t - (D_4 + C_2) \cdot [WC\text{-}1]_t \tag{2.5}$$

$$\frac{d[frq^{r1}]_t}{dt} = S_3 \cdot (1 - g_{f,k}) + S_4 \cdot g_{f,k} - D_3 \cdot [frq^{r1}]_t \tag{2.6}$$

$$\frac{d[FRQ]_t}{dt} = L_3 \cdot [frq^{r1}]_t - D_6 \cdot [FRQ]_t \tag{2.7}$$

$$\frac{d[WCC]_t}{dt} = C_2 \cdot [WC\text{-}1]_t - (D_8 + P \cdot [FRQ]_t{}^{m_{\mathrm{F}}}) \cdot [WCC]_t \tag{2.8}$$

$$\frac{d[ccg^{r1}]_t}{dt} = S_c \cdot g_{g,k} - D_{cr} \cdot [ccg^{r1}]_t \tag{2.9}$$

$$\frac{d[CCG]_t}{dt} = L_c \cdot [ccg^{r1}]_t - D_{cp} \cdot [CCG]_t \tag{2.10}$$

$$\frac{d\Psi(t \mid t_0 = t_k^{(\mathrm{gf})})}{dt} = \sum_{\mathrm{x} \in \{\mathrm{f,c}\}} \left(\delta_{g_{\mathrm{x},k},0}\ A_{\mathrm{x}}([WCC]_t)^{n_{\mathrm{W}}} + \delta_{g_{\mathrm{x},k},1}\ B_{\mathrm{x}} \right) \tag{2.11}$$

Here, $g_{f,k}$ and $g_{c,k}$ are kept constant at the values assigned to them at the most recent gene flipping event, at time $t_k^{(gf)}$. Likewise, the initial values for the continuous non-gene species concentrations at $t = t_k^{(gf)}$ are given by their final values, obtained by ODE solution during the preceding time stepping interval, from $t_{k-1}^{(gf)}$ to $t_k^{(gf)}$, if $k \geq 1$.

At the beginning of the very first time step, at $t = t_0^{(gf)} = 0$, all initial values must be user-supplied: they are given in Table 2.2 below. From Equation 2.1 follows that the initial value of $\Psi(t, t_k^{(gf)})$ must always be reset to zero at the beginning of every new τ step, i.e., $\Psi(t = t_k^{(gf)} + 0^+, t_k^{(gf)}) = 0$ for all $k \geq 0$.

(iii) Once a solution for τ has been found, to satisfy Eq. 2.1 for given u, one of the two genes, either $x = f$ or $x = c$, must be randomly chosen to be flipped at time $t_{k+1}^{(gf)} = t_k^{(gf)} + \tau$. For the gene that *is* chosen to be flipped, we replace its current value, $g_{x,k}$, by its new value $g_{x,k+1} = 1 - g_{x,k}$. For the other gene, *not* chosen to be flipped, we set $g_{x,k+1} = g_{x,k}$. The probability for choosing gene x to be flipped is given by:

$$p_{\mathrm{x},k+1}^{(\mathrm{gf})} = \Gamma_{\mathrm{x}}(g_{\mathrm{x},k}, t_{k+1}^{(\mathrm{gf})}) / \sum_{\mathrm{x}' \in \{\mathrm{f},\mathrm{c}\}} \Gamma_{\mathrm{x}'}(g_{\mathrm{x}',k}, t_{k+1}^{(\mathrm{gf})}) \ . \tag{2.12}$$

To make this random gene choice in our algorithm, we draw another uniform random number, w, from (0,1]. If $w < p_{x,k+1}^{(gf)}$ for $x = f$, we flip the *frq*-gene; else, we flip *ccg*-gene. After flipping the chosen gene, a new cycle of the time stepping process shall be started by randomly choosing τ in the first step.

To derive the Gillespie criterion, Eq. 2.1 and Eq. 2.11 above, consider the conditional probability, $dP(\tau, x|g_{x,k})$, for flipping gene $x = frq$ or $x = ccg$ during infinitesimal time interval $[t, t + d\tau]$ with $t := t_k^{(gf)} + \tau$, given that $g_{x,k}$ has not flipped and kept its value constant until time t, since last gene flip occurred at $t_k^{(gf)}$. This conditional gene flipping probability is determined by the WCC-binding or WCC-unbinding rate, given by:

(i) $A_x \cdot ([WCC]_t)^{n_{\mathrm{w}}}$, for $g_{x,k} = 0 \rightarrow g_{x,k+1} = 1$ (WCC-binding);

(ii) B_x, for $g_{x,k} = 1 \rightarrow g_{x,k+1} = 0$ (WCC-unbinding).

Namely:

$$dP(\tau, x|g_{x,k}) = \Big(A_x([WCC]_t)^{n_W} \delta_{g_{x,k},0} + B_x \delta_{g_{x,k},1} \Big) d\tau \tag{2.13}$$

or, for short:

$$dP(\tau, x|g_{x,k}) = \Gamma_x(g_{x,k}, t)d\tau \tag{2.14}$$

where Γ_x is defined in your Eq. 2.2.

Hence, the cumulative probability for flipping one of the genes, $x = f = frq$ or $x = c = ccg$, during $[t_k^{(gf)}, t_k^{(gf)} + \tau]$ is

$$P(\tau) = 1 - exp[-\int_0^\tau \sum_x \Gamma_x(g_{x,k}, t_k^{(gf)} + \tau')d\tau'] \tag{2.15}$$

Also, the probability density function (PDF) for the flipping one of the genes at time, $t = t_k^{(gf)} + \tau$, is then given by

$$p(\tau) = dP(\tau)/d\tau. \tag{2.16}$$

To numerically generate a random τ-value that is drawn from any PDF $p(\tau)$, we simply have to use the corresponding cumulative probability, i.e. $P(\tau)$ to solve the following equation for τ:

$$P(\tau) = u' \tag{2.17}$$

where u' is a uniform random number, drawn from the interval (0,1). By general theorem, we are then guaranteed that τ is randomly drawn from, i.e. is randomly distributed according to, the PDF $p(\tau)$.

The foregoing Eq. 2.17 can be re-written as

$$\int_0^\tau \sum_x \Gamma_x(g_{x,k}, t_k^{(gf)} + \tau')d\tau' = -ln(1 - u') \tag{2.18}$$

Now change integration variable from τ' to $t' = t_k^{(gf)} + \tau'$ and set $u := 1 - u'$. Notice that u is a uniform random number, drawn from the interval (0,1), since u' is. With these two substitutions, we then get Eq. 2.1, i.e.

$$\int_{t_k^{(gf)}}^{t_k^{(gf)}+\tau} \sum_x \Gamma_x(g_{x,k}, t')dt' = -ln(u) \tag{2.19}$$

Once the random τ has been chosen by Eq. 2.1, the probability for flipping gene x is then given by:

$$p_{x,k+1}^{(gf)} = dP(\tau, x|g_{x,k})/\Big(\sum_{x'} dP(\tau, x'|g_{x',k})\Big) \tag{2.20}$$

Using $t_k^{(gf)} + \tau = t_{k+1}^{(gf)}$ and using Eq. 2.14 above, this can then be written in the form of Eq. (2.12), i.e.,

$$p_{x,k+1}^{(gf)} = \Gamma_x(g_{x,k}, t_{k+1}^{(gf)})/\Big(\sum_{x'} \Gamma'_x(g_{x',k}, t_{k+1}^{(gf)})\Big) \tag{2.21}$$

In conclusion, randomness mainly happens at two places in the hybrid model. One is at the place where we want to choose the value of the time step, τ, by using Eq. 2.1. The other one is at the place where a gene is randomly picked to be flipped with the probability in Eq. 2.12. These two parts in the hybrid model are not only the major source of the stochasticity, but, more importantly, the ultimate driver of the clock oscillations.

2.2 Simulation results of the single-cell hybrid model

Apart from all the kinetic rate equations and the KMC process for the system, effective reaction rate parameters and species's initial conditions are highly essential to perform the simulation of the hybrid model as well. Table 2.1 and 2.2 provide the values and units of all rate coefficients and species' initial conditions for the hybrid model simulation in the dark [15]. All concentrations in the simulation are in units of number of molecules per cell (NPC). The amounts of cooperativities in reaction P, A_f, and A_c, namely $\mathrm{n_W}$ and $\mathrm{m_F} = 4$, are from [11].

Table 2.1: Values for reaction rate parameters of the hybrid model. All concentrations in the simulation are in units of number of molecules per cell (NPC).

Parameter	Value	Unit	Parameter	Value	Unit
A_f	2.560E-10	$NPC^{-n_{\mathrm{W}}} \cdot h^{-1}$	D_6	0.1518	h^{-1}
B_f	1.590	h^{-1}	D_7	0.1384	h^{-1}
S_1	83.71	$NPC \cdot h^{-1}$	D_8	2.487E-3	h^{-1}
S_3	3.569	h^{-1}	C_2	0.1627	h^{-1}
S_4	5453	h^{-1}	P	3.120E-11	$NPC^{-m_{\mathrm{F}}} \cdot h^{-1}$
D_1	0.7237	h^{-1}	A_c	1.860E-08	$NPC^{-n_{\mathrm{W}}} \cdot h^{-1}$
D_3	0.2997	h^{-1}	B_c	2.581	h^{-1}
C_1	4.810E-05	$NPC^{-1} \cdot h^{-1}$	S_c	73.80	h^{-1}
L_1	4.245	h^{-1}	L_c	2.231	h^{-1}
L_3	0.4851	h^{-1}	D_{cr}	0.2198	h^{-1}
D_4	3.233E-3	h^{-1}	D_{cp}	0.6969	h^{-1}
n_{W}	4		m_{F}	4	

Table 2.2: Initial concentrations for all the species in the hybrid model. All concentrations in the simulation are in units of number of molecules per cell (NPC).

Species	Initial Concentration (NPC)
$wc\text{-}1^{r0}$	113
$wc\text{-}1^{r1}$	18
$WC\text{-}1$	459
frq^{r1}	31
FRQ	345
WCC	101
ccg^{r1}	26
CCG	102
$g_{f,k=0}$	0
$g_{c,k=0}$	0

Programs for all simulation works in this dissertation are written in C++ for efficiency. To generate the uniform random numbers, u and w, required for generating τ and choosing the gene to be flipped, the random number generator, *mt19937* [16], is used as the major randomization source. To solve the rate equation system, Eqs. 2.4-2.11, I used the classic Runge-Kutta method [17], also known as 4th order Runge-Kutta method, as the ODE solver, where the simulation time step size of the ODE solver is $\tau_{sim} = 0.01h$, and the observation time step size of the output time series is $\tau_{obs} = 0.5h$.

To estimate the rough systematic numerical error in the ODE solver, due to the simulation step size, $\tau_{sim} = 0.01h$, we ran the single-cell hybrid model twice for $T_{test} = 3h$, within which the gene flip event has not yet happened, with the same initial conditions and parameters but different simulation time step sizes, $\tau_{1,sim} = 0.01h$ and $\tau_{2,sim} = 0.005h$. Then we compared their WCC concentration time series, $[WCC]_i^{(\mathrm{run1})}$ and $[WCC]_j^{(\mathrm{run2})}$, corresponding to $\tau_{1,sim}$ and $\tau_{2,sim}$, respectively, by using the mean percentage difference,

$$\overline{diff\%} = \sum_{i=1}^{N_{test}} \frac{[WCC]_i^{(\mathrm{run1})} - [WCC]_{2i}^{(\mathrm{run2})}}{([WCC]_i^{(\mathrm{run1})} + [WCC]_{2i}^{(\mathrm{run2})})/2} \tag{2.22}$$

where $N_{test} = T_{test}/\tau_{1,sim}$. The mean percentage difference for systems with $\tau_{1,sim} = 0.01h$ and $\tau_{2,sim} = 0.005h$ is $\overline{diff\%} = 3 \times 10^{-11}\%$, which indicates that the simulation step width, $\tau_{1,sim} = 0.01h$, is small enough to get accurate results.

To generate a typical trajectory sample consisting of $N_R = 1000$ single-cell trajectories over a 240h physical simulation time interval, it takes about four minutes of single-processor CPU time on the processor of AMD Ryzen 7 3700X, using the parameter values given in Tables 2.1 and 2.2 below. The CPU time scales roughly linearly with the number of single-cell trajectories and with the length of the physical simulation time interval.

In the simulation of the hybrid model, a specific time position for $(k + 1)$-th gene flip event should be located by estimating the time when the Gillespie Criterion, Eq. 2.1, is met.

Due to the discrete time step in the ODE solver, we can directly compare the integration result, $\Psi(t|t_0 = t_k^{(\mathrm{gf})})$, of Equation 2.11 with $-ln(u_k)$, where the random number, u_k, is generated right after the most recent gene flip event at $t_k^{(\mathrm{gf})}$, at each simulation time step by employing a binary search-type [18] method, as follows:

(i) We start right after the k-th gene flip event at $t_k^{(\mathrm{gf})}$, which happens within the i-th simulation time step with a time interval $[t_i, t_{i+1}]$, where $t_{i+1} = t_i + \tau_{sim}$ and $\tau_{sim} = 0.01h$.

(ii) After solving the ODE solver for the time interval $[t_k^{(\mathrm{gf})}, t_{i+1}]$, if $\Psi(t_{i+1}|t_0 = t_k^{(\mathrm{gf})}) < -ln(u_k)$, it means that the Gillespie criterion is not yet met, and the ODE solver can continue to the $(i+1)$-th simulation time step without changing $g_{x,k}$.

(iii) When the Gillespie criterion is met, $\Psi(t_{i+l}|t_0 = t_k^{(\mathrm{gf})}) \geq -ln(u_k)$, it indicates that the $(k+1)$-th gene flip event happens within the $(i+l-1)$-th simulation time step, and we set $t_{left} = t_{i+l-1}$ or $t_k^{(\mathrm{gf})}$ $(l = 1)$ and $t_{right} = t_{i+l}$ as the initial boundaries for the binary search-type method.

(iv) Let $t_{mid} = 0.5 \cdot (t_{left} + t_{right})$, and use the ODE solver to estimate $\Psi(t_{mid}|t_0 = t_k^{(\mathrm{gf})})$. If $\Psi(t_{mid}|t_0 = t_k^{(\mathrm{gf})}) > -ln(u)$, set $t_{right} = t_{mid}$; otherwise, set $t_{left} = t_{mid}$.

(v) Repeat (iv) to find $t_{k+1}^{(\mathrm{gf})}$ until the system reaches either of the two converging thresholds, $|t_{left} - t_{right}| \leq 10^{-5} \cdot (t_{i+1} - t_i)$ or $|\Psi(t_{mid}|t_0 = t_k^{(\mathrm{gf})}) + ln(u)| \leq 10^{-5} \cdot |\Psi(t_{i+1}|t_0 = t_k^{(\mathrm{gf})}) - \Psi(t_i|t_0 = t_k^{(\mathrm{gf})})|$.

After pinpointing the $(k+1)$-th gene flip event, we can flip the gene randomly selected by Equation 2.12.

Figure 2.2 shows simulation results of the concentration time series for four different species in one cell. Smooth but aperiodic oscillations can be observed for each species in Figure 2.2. If we start the cycle at the time where a black arrow points, the first thing to be seen is that the concentration time series of frq^{r1}, $[frq^{r1}]_t$, rises right after the black arrow. $[FRQ]_t$ follows the trace of $[frq^{r1}]_t$ because the production of FRQ is only affected by $[frq^{r1}]_t$. Then, due to the degrading of frq^{r1} and FRQ, $[WCC]_t$ start to rise after some point, where consumption rate of WCC in reaction P is slower than the generation rate of WCC in reaction C_2. When $[WCC]_t$ rises, the random threshold of Ψ in Equation 2.11 will be finally reached. As a result, frq-gene is, again, flipped to its active state, which brings the system back to the start place of the cycle. In conclusion, we can see that the random gene flip events drive the clock oscillations in the system.

In the following, I will frequently make use of power spectra, denoted $S(f)$, to analyze time species concentration series data, $[X]_t$, that generated by the generalized Gillespie algorithm described above. For a given physical simulation time interval $[t_{ini}, t_{fin}]$, these data are given on a dense, discrete simulation time grid with grid points, $t_j^{(s)} = (j-1) \cdot \Delta t^{(s)} + t_{ini}$ for $j = 1, 2, ..., J+1$, where $J+1$ is the number of grid points and $\Delta t^{(s)} = (t_{fin} – t_{ini})/J$. We take the discrete Fourier transform [19] of $[X]_t$ on the time grid $t_j^{(s)}$,

$$\hat{X}(f_\nu) = \frac{1}{J+1} \sum_{j=1}^{J+1} exp(-2\pi i f_\nu t_j^{(s)}) \cdot [X]_t \tag{2.23}$$

with grid of allowed frequencies is $f_\nu = \nu/(t_{fin} – t_{ini})$ for $\nu = 0, 1, ...(J+1)/2$. We then obtain the power spectrum, $S(f_\nu)$, by

$$S(f_\nu) = |\hat{X}(f_\nu)|^2 \tag{2.24}$$

Power spectra of the four species' time series trajectories are displayed in Figure 2.3. Means and error bars are obtained by using the Bootstrap sampling method [20]:

(i) Generate a base sample of $N_R = 100$ different random trajectories from the hybrid model, by the Gillespie KMC algorithm described above, using 100 different random number sequences.

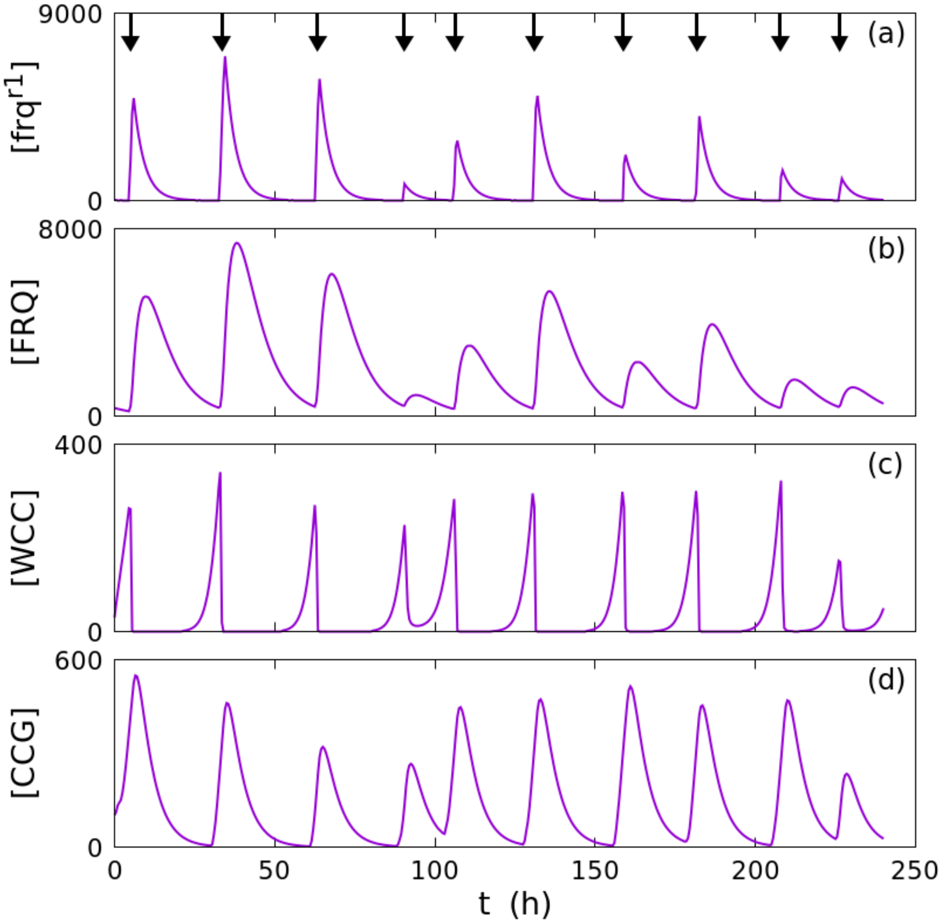

Figure 2.2: Simulation results of the single-cell hybrid model in the dark for 10 days. All signals in (a)-(d) are obtained from only one representative, randomly chosen single-cell result. (a) frq mRNA concentration trajectory. Black arrows indicates the time, which has an average duration of $1/B_f = 0.629h$, when frq gene is in active state. (b) FRQ protein concentration trajectory. (c) WCC protein concentration trajectory. (d) CCG protein concentration trajectory.

(ii) To generate a single bootstrap sample, randomly draw $N_R = 100$ trajectories, with replacement, from the base sample of $N_R = 100$ trajectories generated in (i). To do so, draw N_R uniformly distributed random integer numbers, n_1, ..., n_{N_R}, from the interval [1, 100]. From the base sample, include in the bootstrap sample the trajectories that are indexed n_1 though n_R in the base sample. In this process, some trajectories may be included multiple times in the bootstrap sample, as many times their index, n, appears in the random list n_1, ..., n_{N_R}.

(iii) For each trajectory included in the bootstrap sample from (ii), calculate the power spectrum, $S(f)$, by Eqs. 2.23 and 2.24, from the concentration time series, $[X]_t$, for the species of interest, X. Then average $S(f)$ over all trajectories in the bootstrap sample and denote the result by $\overline{S(f)}$.

(iv) Repeat (ii)-(iii) for $B = 1000$ times with the power spectrum mean, $\overline{S_b(f)}$, for each bootstrap sample b, where $b = 1, 2, 3, ..., B$. Then calculate the mean and standard error of the mean (SEM), $\overline{S_b(f)}$, over $B = 1000$ bootstrap samples.

As shown in Figure 2.3, a sharp peak, indicating the intrinsic period of the clock, can be observed in each power spectrum curve, together with the major period labeled next to it. The major period of the single clock oscillator estimated by the hybrid model is within the range from 24h to 27h, whereas the experiment results show that the circadian period of *N. crassa* is within the range from 22h to 23h [7, 21–23]. There are two possible factors of this period difference between simulation and experiment results: (i) the default initial conditions and parameter set used in the single-cell hybrid model are not specifically obtained from fitting experiment data to the single-cell hybrid model; (ii) the experiment results of 22-23h period are mainly obtained from strains of fungi, while the simulation result is only for single isolated cells.

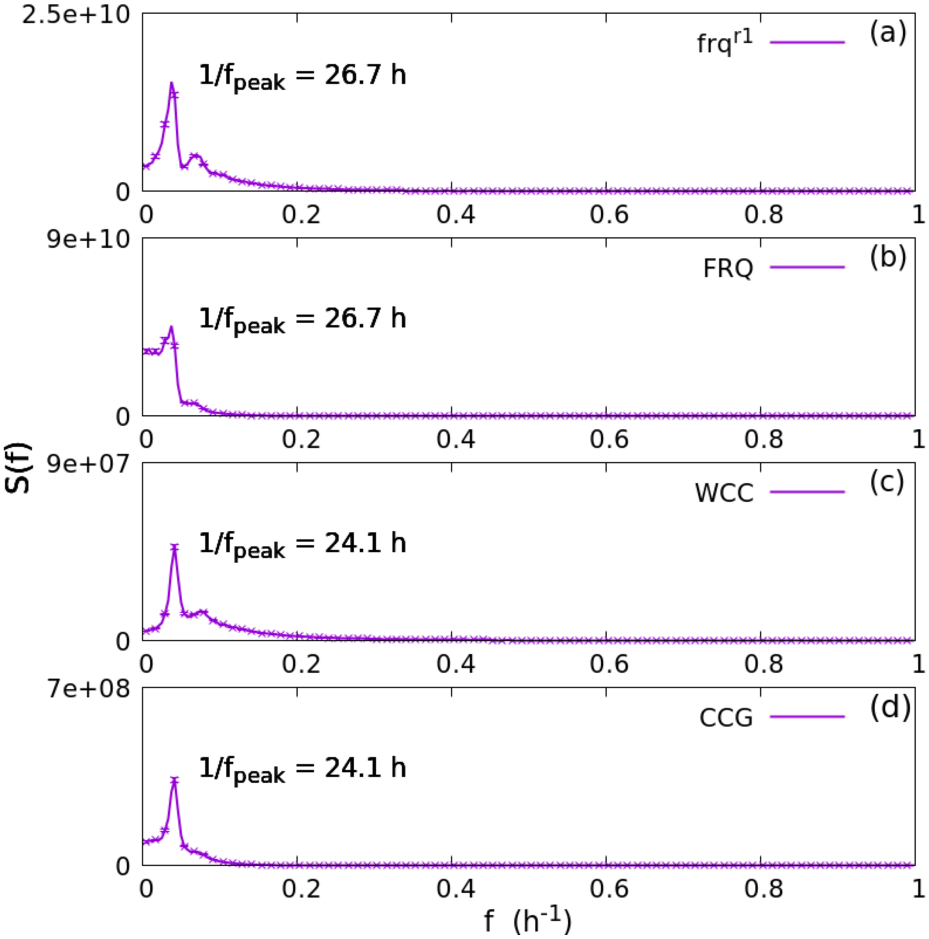

Figure 2.3: Power spectra of the concentration trajectories, with $(t_{ini} = 0h, t_{fin} = 239.5h, J = 480)$, for four species, frq^{r1}, FRQ, WCC, and CCG . Means and error bars with ±1 standard error of the mean (SEM) are obtained by bootstrapping a sample of $N_{\mathrm{R}} = 1000$ cell signals and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

2.3 Single-cell hybrid model with light entrainment

Light exposure is one of the essential factors that regulates the circadian rhythms of Neurospora crassa [4–6]. To introduce the feature of responsiveness to visible light stimuli into the hybrid model, a reaction C_3, in which the production of WCC can be affected by the interactions between photons and WC-1, is added to the network [24].

Figure 2.4 displays the genetic network of hybrid model with the light coupling part. Because reaction C_3 only relates to WC-1 and WCC, we can include it in Equation 2.5 and 2.8:

$$\frac{d[WC\text{-}1]_t}{dt} = L_1 \cdot [wc\text{-}1^{r1}]_t - (D_4 + C_2) \cdot [WC\text{-}1]_t - Phot \cdot C_3 \cdot [WC\text{-}1]_t \tag{2.25}$$

$$\frac{d[WCC]_t}{dt} = C_2 \cdot [WC\text{-}1]_t - (D_8 + P \cdot [FRQ]_t{}^{m_F}) \cdot [WCC]_t + Phot \cdot C_3 \cdot [WC\text{-}1]_t \tag{2.26}$$

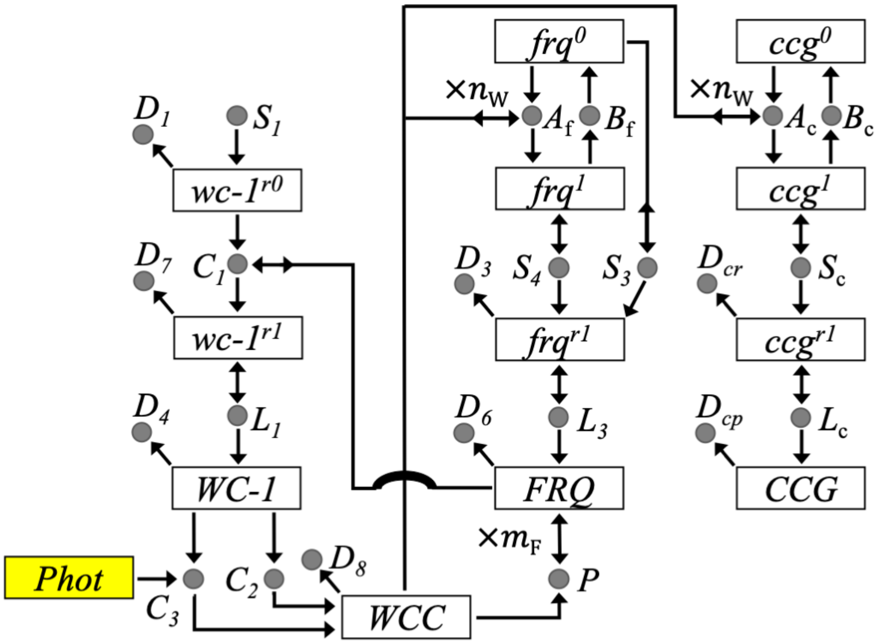

Figure 2.4: A simplified genetic network for the biological clock of Neurospora crassa with light entrainment. The yellow box denotes a photon species.

$Phot$ in Equation 2.25 and 2.26 is a binary variable, with allowed values, $Phot = 1$, if light is on, and $Phot = 0$, if light is off. Based on this modification to the hybrid model, the system can now be easily entrained with artificial days of different light exposure periods, like an 6+6h artificial day, which represents 6 hours with light exposure and 6 hours in the dark. To determine the light status of the system in the simulation, we can simply use the equation, shown below, at the beginning of each simulation time step.

$$Phot = (\lfloor t_i/t_{light} \rfloor + Phot_0) \bmod 2 \tag{2.27}$$

where t_i is the time step of the ODE solver, t_{light} is the light exposure period for each artificial day, and $Phot_0$ indicates the starting light status of the simulation.

Figure 2.5 shows WCC's concentration time series, $[WCC]_t$, of hybrid models entrained with three artificial days, 6+6h, 12+12h, and 18+18h, while Figure 2.6 shows the power spectra of WCC corresponding to those three artificial days. For simulations in both Figure 2.5 and Figure 2.6, the reaction rate C_3 is equal to $2 \cdot C_2$, and the reaction rate parameters and initial conditions are from Table 2.1 and 2.2. Means and error bars in Figure 2.6 and Figure 2.7 are obtained from bootstrap method with a sample of $N_R = 1000$ system trajectories and $B = 1000$ bootstrap samples. The phasing of the light exposure for all simulations in the section are L/D, which means that the system starts with light in a cycle, and then in the dark for the other half of artificial day period later.

$[WCC]_t$ of three light exposure periods, in Figure 2.5, exhibit irregular oscillations, even though each system is subject to the light exposure of a certain period. This result is confirmed by the corresponding power spectra in Figure 2.6, where a dominant approximate 24-hour peak is exhibited in the each power spectrum curve despite the light exposure. In Figure 2.6, frequencies caused by the light exposures can be easily found next to or on the

24-h peak in panels (a) and (b), whereas the clock oscillator does not display very pronounced response to the 18+18h light exposure in panel (c).

Figure 2.7 shows the power spectra of WCC's concentration with different C_3 values, $C_3 = 0.5 \cdot C_2$, C_2, and $2.0 \cdot C_2$. It is obvious to see that the power spectrum with $C_3 = 2 \cdot C_2$ have a larger peak at frequency, $f = 1/12h^{-1}$, than those with smaller values of C_3, which indicates that the light strength is well embedded in the reaction C_3 and can be easily adjusted by changing the C_3 value.

In summary, the simulation results show that our model does have response to light stimuli, but with a limited extent. The randomness, which comes from the gene flip events, still dominates the clock oscillations in the single-cell system, despite strong light exposures to which the system is subject.

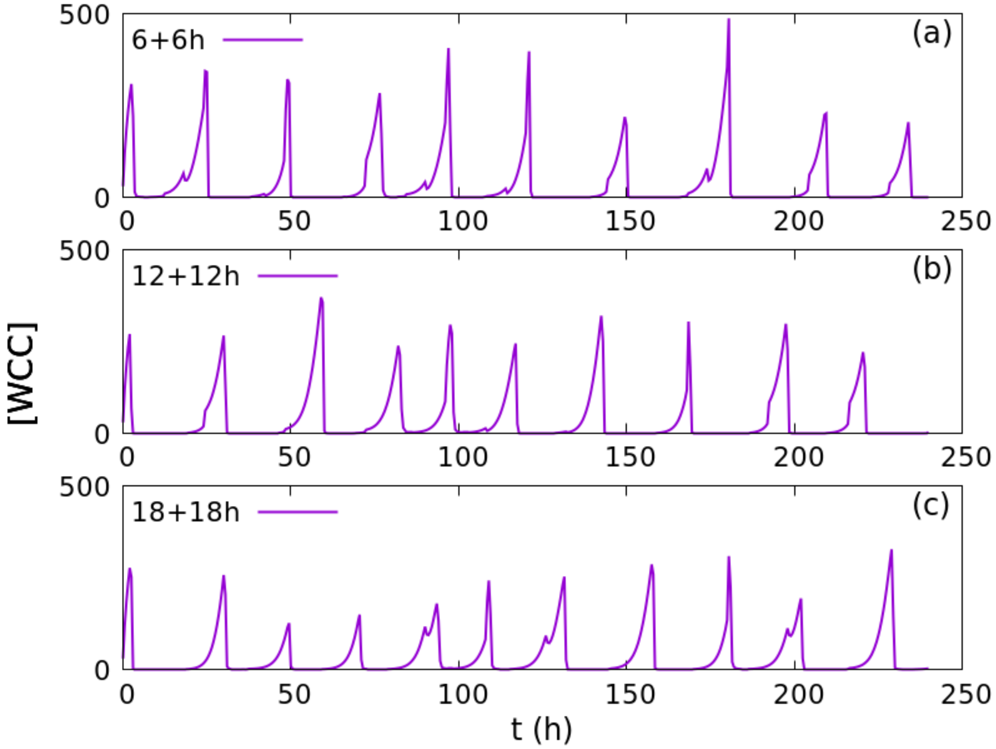

Figure 2.5: $[WCC]$ trajectories entrained with different light-dark periods. Each signal in (a)-(c) is obtained from only one representative, randomly chosen single-cell result. (a) Light exposure period is 6+6h. (b) Light exposure period is 12+12h. (c) Light exposure period is 18+18h.

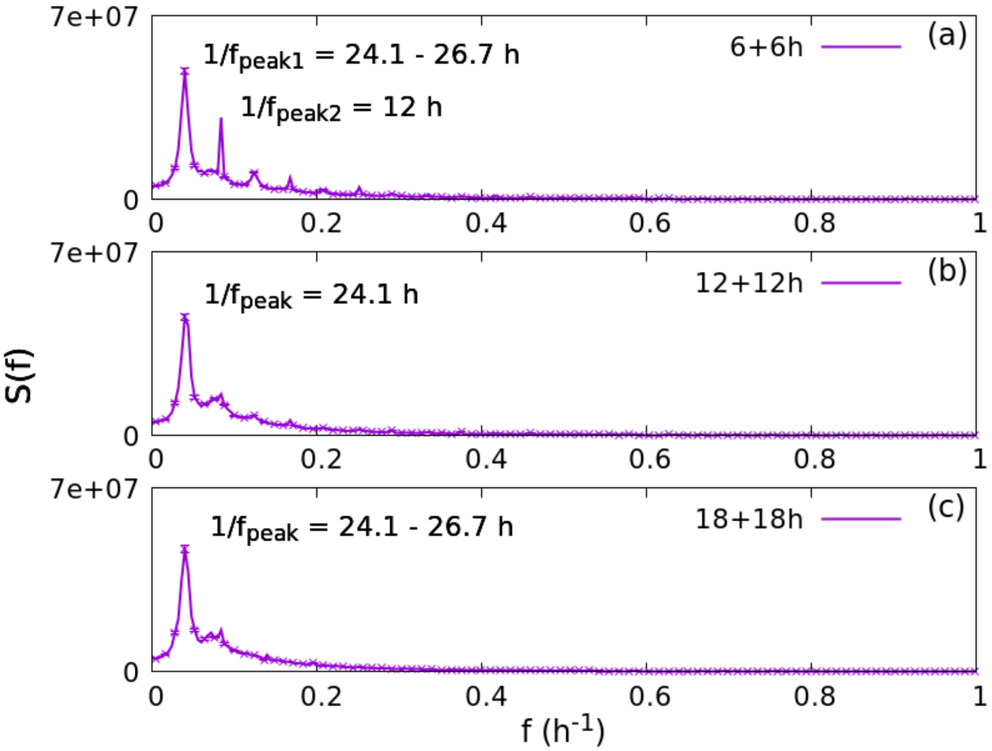

Figure 2.6: Power spectra of $[WCC]$ signals, with $(t_{ini} = 0h, t_{fin} = 239.5h, J = 480)$, subject to three artificial days, 6+6h, 12+12h, and 18+18h. Means and ±1 SEM error bars are obtained from bootstrap sampling method from a sample of $N_{\mathrm{R}} = 1000$ single-cell trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

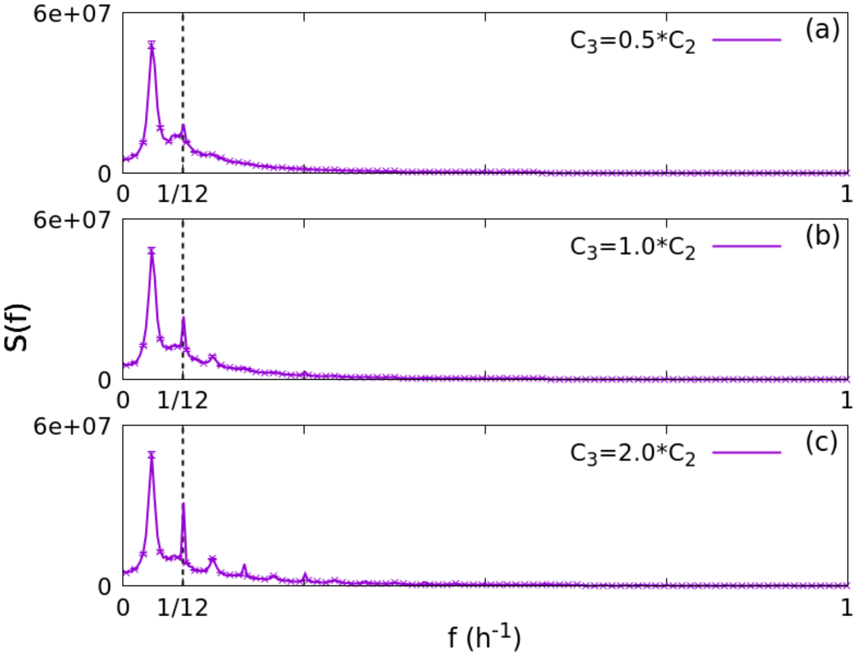

Figure 2.7: Power spectra of $[WCC]$ trajectories, with $(t_{ini} = 0h, t_{fin} = 239.5h, J = 480)$, for different C_3 values. The light exposure period is 6+6h. Means and ±1 SEM error bars are obtained from the bootstrap sampling method from a sample of $N_{\mathrm{R}} = 1000$ single-cell trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot. (a) $C_3 = 0.5 \cdot C_2$. (b) $C_3 = 1.0 \cdot C_2$. (c) $C_3 = 2.0 \cdot C_2$.

Chapter 3

Hybrid model combined with a quorum sensing coupling mechanism for multi-cell system

3.1 The multi-cell hybrid model with quorum sensing

To study the clock synchronization in a population of stochastic coupled single-cell oscillators, the hybrid model is then expanded to the multi-cellular level, with a quorum sensing-type signaling interaction between cells [25, 26]. Fig. 3.1 presents the network of this multi-cell hybrid model in dark.

A chemical signaling molecule, Si, is assumed to be produced in each cell by the protein product of gene *ccg*. This signaling molecule is then both exported to and imported from an extra-cellular medium, shared by all cells in the population, with a diffusion rate η. Intra-cellular Si modulates clock oscillations by either suppressing or enhancing the rate of WCC production in the reaction C_2 and C_4. To include this modulation in the kinetic rate equations, the rate coefficient, C_2, in both Equation 2.5 and 2.8 now is changed to C_{eff} with

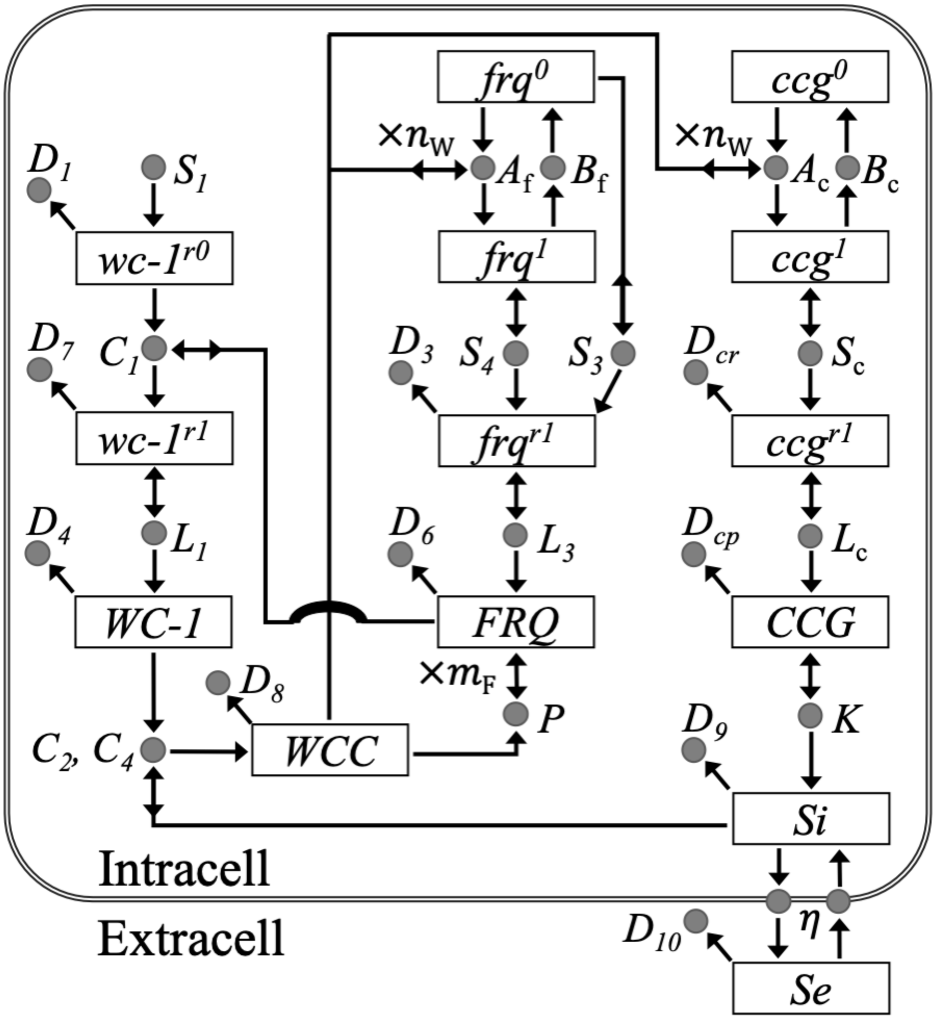

Figure 3.1: Multi-cell hybrid model

the equation,

$$C_{eff}([Si]_{n,t}) = X \cdot H(X), \quad H(X) = \begin{cases} 1 & \text{if } X \geq 0 \\ 0 & \text{if } X < 0 \end{cases} \tag{3.1}$$

where $X := C_2 - C_4 \cdot [Si]_{n,t}$, and $[Si]_{n,t}$ is the Si's concentration in cell n at time t.

Basically, C_{eff} is the combination of the reaction C_2 and C_4. If Si is greater than or equal to C_2/C_4, the production of protein WCC would be immediately stopped. This gives another negative-feedback element to the system to control the production of WCC, and then to adjust the oscillation phase of the clock.

In the inter-cellular medium, the concentration of the signaling molecule, Se, is mainly affected by degradation and the diffusion towards/from cells. Its rate equation is given below [25],

$$\frac{d[Se]_t}{dt} = -D_{10} \cdot [Se]_t + \eta_{ext} \cdot \sum_{n=1}^{N} ([Si]_{n,t} - [Se]_t), \tag{3.2}$$

where η_{ext} is the diffusion rate of Se across the cell membrane. Because all cells in the population share the same medium, $[Se]_t$ does not need to have a subscript, n, to identify which cell it belongs to.

Based on the quasi-steady-state approximation [25,27,28], we can derive the approximation of Se concentration at each time step, which is shown below [25], from simply setting its rate equation, Equation 3.2, to zero.

$$[Se]_t = Q \cdot \frac{1}{N} \sum_{n}^{N} [Si]_{n,t} \tag{3.3}$$

where Q [25] is

$$Q = \frac{N \cdot \eta_{ext}}{D_{10} + N \cdot \eta_{ext}} \tag{3.4}$$

The diffusion rate of Se across the membrane has the form, $\eta_{ext} = \sigma \cdot A_{surface}/V_{ext}$ [25], with σ being the membrane permeability, $A_{surface}$ being the cell surface area, and V_{ext} being the inter-cellular medium volume. Based on the definition of Se's diffusion rate, Equation 3.4 indicates that Q is a monotonic function of the cell density, N/V_{ext}, within the range of [0,1]. As a result, we can directly use Q instead of cell density to control the coupling strength of cells in the model.

Due to the expansion of the model to describe a system of multiple cells, we put a label 'n' in the subscript of each species' concentration variable to indicate which cell it belongs to, and integrate all cells' effective rate coefficients for flipping genes' states. Therefore, we now use Eq. 3.5 to choose the random τ-step between successive gene flips by using the effective rate coefficient defined in Eq. 3.6, and then use Eq. 3.7 to choose a random cell, n, and a random gene x, to be flipped in that cell n.

$$\Psi(t = t_{k+1}^{(\mathrm{gf})} \mid t_0 = t_k^{(\mathrm{gf})}) \equiv \int_{t_k^{(\mathrm{gf})}}^{t_k^{(\mathrm{gf})}+\tau} dt' \sum_{\mathrm{x}\in\{\mathrm{f,c}\}} \sum_{n=1}^{N} \Gamma_{\mathrm{x},n}(g_{\mathrm{x},n,k}, t') = -\ln(u) \ , \tag{3.5}$$

$$\Gamma_{\mathrm{x},n}(g_{\mathrm{x},n,k}, t) \ = \ \delta_{g_{\mathrm{x},n,k},0} \ A_{\mathrm{x}}([WCC]_t)^{n_{\mathrm{W}}} \ + \ \delta_{g_{\mathrm{x},n,k},1} \ B_{\mathrm{x}} \ , \tag{3.6}$$

$$p_{\mathrm{x},n,k+1}^{(\mathrm{gf})} = \Gamma_{\mathrm{x},n}(g_{\mathrm{x},n,k}, t_{k+1}^{(\mathrm{gf})}) / \sum_{\mathrm{x}'\in\{\mathrm{f,c}\}} \sum_{n'=1}^{N} \Gamma_{\mathrm{x}',n'}(g_{\mathrm{x}',n',k}, t_{k+1}^{(\mathrm{gf})}) \ . \tag{3.7}$$

The kinetic rate equations of the multi-cell hybrid model are given as follows,

$$\frac{d[wc\text{-}1^{r0}]_{n,t}}{dt} = S_1 - D_1 \cdot [wc\text{-}1^{r0}]_{n,t} - C_1 \cdot [wc\text{-}1^{r0}]_{n,t} \cdot [FRQ]_{n,t} \tag{3.8}$$

$$\frac{d[wc\text{-}1^{r1}]_{n,t}}{dt} = C_1 \cdot [wc\text{-}1^{r0}]_{n,t} \cdot [FRQ]_{n,t} - D_7 \cdot [wc\text{-}1^{r1}]_{n,t} \tag{3.9}$$

$$\frac{d[WC\text{-}1]_{n,t}}{dt} = L_1 \cdot [wc\text{-}1^{r1}]_{n,t} - (D_4 + C_{eff}([Si]_{n,t})) \cdot [WC\text{-}1]_{n,t} \tag{3.10}$$

$$\frac{d[frq^{r1}]_{n,t}}{dt} = S_3 \cdot (1 - g_{f,n,k}) + S_4 \cdot g_{f,n,k} - D_3 \cdot [frq^{r1}]_{n,t} \tag{3.11}$$

$$\frac{d[FRQ]_{n,t}}{dt} = L_3 \cdot [frq^{r1}]_{n,t} - D_6 \cdot [FRQ]_{n,t} \tag{3.12}$$

$$\begin{aligned}\frac{d[WCC]_{n,t}}{dt} &= C_{eff}([Si]_{n,t}) \cdot [WC\text{-}1]_{n,t} - D_8 \cdot [WCC]_{n,t} \\ &\quad - P \cdot [FRQ]_{n,t}{}^{m_F} \cdot [WCC]_{n,t}\end{aligned} \tag{3.13}$$

$$\frac{d[ccg^{r1}]_{n,t}}{dt} = S_c \cdot g_{g,n,k} - D_{cr} \cdot [ccg^{r1}]_{n,t} \tag{3.14}$$

$$\frac{d[CCG]_{n,t}}{dt} = L_c \cdot [ccg^{r1}]_{n,t} - D_{cp} \cdot [CCG]_{n,t} \tag{3.15}$$

$$\frac{d[Si]_{n,t}}{dt} = K \cdot [CCG]_{n,t} - D_9 \cdot [Si]_{n,t} + \eta \cdot ([Se]_t - [Si]_{n,t}) \tag{3.16}$$

$$\frac{d\Psi(t \mid t_0 = t_k^{(\mathrm{gf})})}{dt} = \sum_{\mathrm{x}\in\{\mathrm{f,c}\}} \sum_{n=1}^{N} \left(\delta_{g_{\mathrm{x},n,k},0} \ A_{\mathrm{x}}([WCC]_{n,t})^{n_{\mathrm{W}}} + \delta_{g_{\mathrm{x},n,k},1} \ B_{\mathrm{x}}\right) \tag{3.17}$$

3.2 Simulation results of the multi-cell system

Before using the multi-cell hybrid model to simulate a population of coupled cells, we assume that all cells share the same kinetics model parameter set and single-cell initial condition (IC) from Table 2.1 and Table 2.2, respectively. For the part of modeling the quorum sensing, all cells have the same IC of $[Si]_{n,t}$, as $[Si]_{n,t=0} = 1.0$, and the same parameter set of quorum sensing model, as shown in Table 3.1.

We use uniform ICs for cells because they are actually closer to the experimental situation, where cell populations are always “synchronized” by strong light exposure, before experimental observations start, at t=0. With uniform ICs, cells completely randomize themselves in very short time, within one diurnal cycle. Therefore, it makes no difference, after one diurnal cycle, whether ICs of cells are randomized across population or uniform.

Table 3.1: Values for the parameters of quorum sensing model. All concentrations in the simulation are in units of number of molecules per cell (NPC).

Parameter	Value	Unit	Parameter	Value	Unit
K	0.05	h^{-1}	C_4	0.9	$(NPC \cdot h)^{-1}$
D_9	26	h^{-1}	η	100	h^{-1}
Q	1.0				

The same random number generator and ODE solver are used as in simulations of the single-cell hybrid model. To generate a typical system trajectory of $N_{=}500$ cells over a 250h physical simulation time interval, it takes about fifteen minutes of single-processor CPU time on the processor of AMD Ryzen 7 3700X, using the default parameter values and uniform ICs given in Tables 2.1-2.2 and 3.1. The CPU time scales roughly quadratic with the number of cells in a population.

Figures 3.2 - 3.5 compare the model simulation results of species FRQ, WCC, CCG, and Si between concentration population average (PA) and intra-cell concentrations. The concentration population average of a species, X, is defined as

$$[X]_{\mathrm{PA,t}} = \frac{1}{N}\sum_{n=1}^{N}[X]_{n,t}, \tag{3.18}$$

where $[X]_{n,t}$ is the intra-cell concentration of the species X. A single random time series of the continuous non-gene concentrations and gene states of all cells in a N-cell population, generated by the multi-cell hybrid model, will be referred to as a "system trajectory" in the following.

In Figures 3.2-3.5, the concentration PA, $[X]_{\mathrm{PA,t}}$, for one random system trajectory of a $N = 500$-cell population is shown in each Panel (a). In the meantime, the intra-cell concentrations, $[X]_{n,t}$, of three random selected cells, which are from the corresponding populations of Panel (a), are shown in Panels (b)-(d). The ×-symbols in Panels (a)-(d) represent the local maxima in the curves. The vertical arrows in Panels (b)-(d) indicate the short frq-gene bursts, where $g_{f,n,k} = 1$, of average duration $1/B_f = 0.63h$, interspersed by transcriptional inactivity, $g_{f,n,k} = 1$. Error bars in Panels (e) of Figures 3.2-3.5 are ±1 standard error of the mean (SEM), generated by the bootstrap sampling method for a random sample of $N_{\mathrm{R}} = 100$ system trajectories.

When we compare PA and single-cell signals, which are belong to the same cell population, in Figures 3.2-3.5, the differences of them are quite obvious, except for $[Si]$. On the one hand, PA signals clearly show coherent and nearly periodic oscillations, but, on the other, single-cell signals, which contribute to the PA signals, have highly irregular oscillations, exhibiting no evidence to any periodicity or synchronization. This interesting contrast also can be observed by comparing the power spectrum of the concentration PA to the PA of the concentration power spectra of individual cells, shown in Panel (e) of Figures 3.2-3.5. The former exhibits a series of strong, sharp peaks, reflecting the nearly periodic, coherent oscillations. The latter only shows a very broad, featureless frequency continuum without any pronounced peak

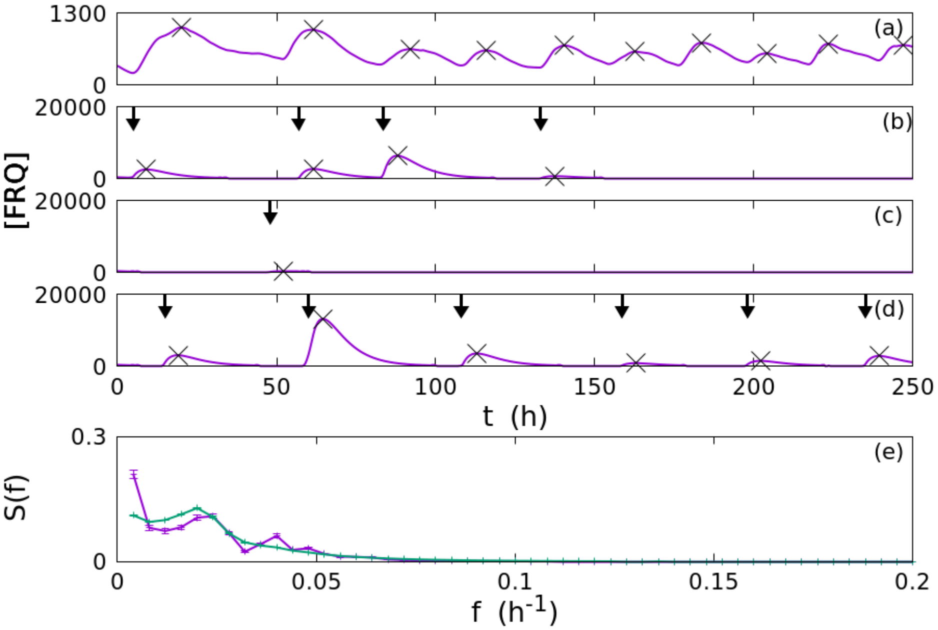

Figure 3.2: Simulation results of multi-cell hybrid model without light exposure. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[FRQ]_{\mathrm{PA,t}}$. (b)-(d) $[FRQ]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[FRQ]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[FRQ]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

features. This again suggests that a large fraction of the individual cells in the population are neither close to oscillating coherently, nor in approximate synchrony with the PA signal.

As mentioned above, the difference between concentration PA, $[Si]_{\mathrm{PA,t}}$ and $[Si]_{n,t}$ of individual cells is much smaller than those of other species. In Figure 3.5, despite the fact that single-cell signals are not so highly coherent and smooth as the PA signal, periodicity can still be observed in their oscillations. Equation 3.16 shows that $[Si]_{n,t}$ mainly comes from two sources: one is the generation reaction controlled by $[CCG]_{n,t}$, and the other is the

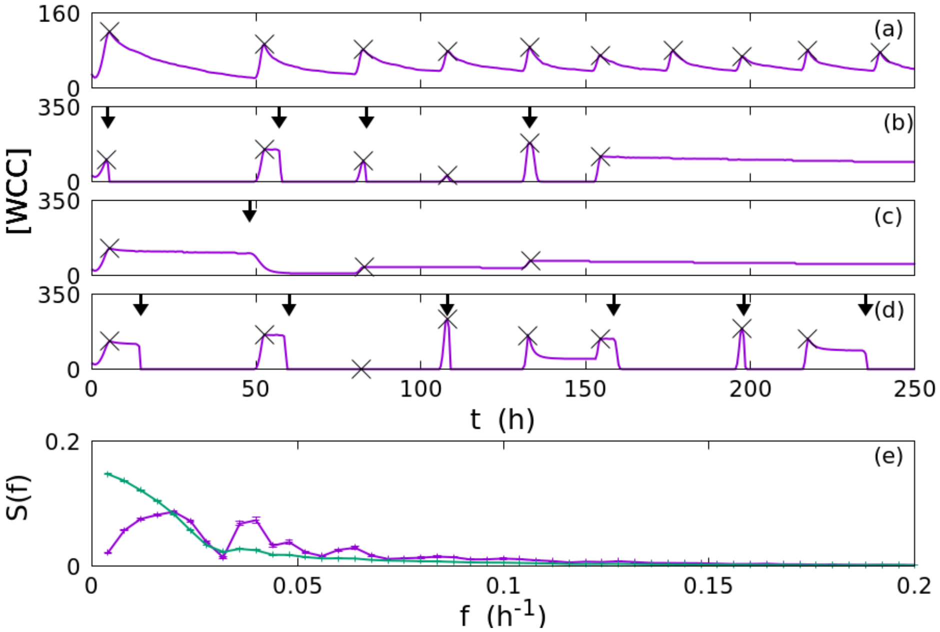

Figure 3.3: Simulation results of multi-cell hybrid model without light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[WCC]_{\mathrm{PA,t}}$. (b)-(d) $[WCC]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[WCC]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[WCC]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

diffusion process controlled by both $[Si]_{n,t}$ and $[Se]_t$, which is equivalent to $[Si]_{\mathrm{PA,t}}$ in our simulation. The former source provides indispensable randomness for $[Si]_{n,t}$, but the latter one is the main factor that sets Si apart from others.

In Figures 3.6-3.11, we compared simulation results of different cell population sizes and signal coupling parameters, Q and C_4. Error bars are ±1 SEM, obtained by Bootstrapping a sample of $N_{\mathrm{R}} = 100$ system trajectories.

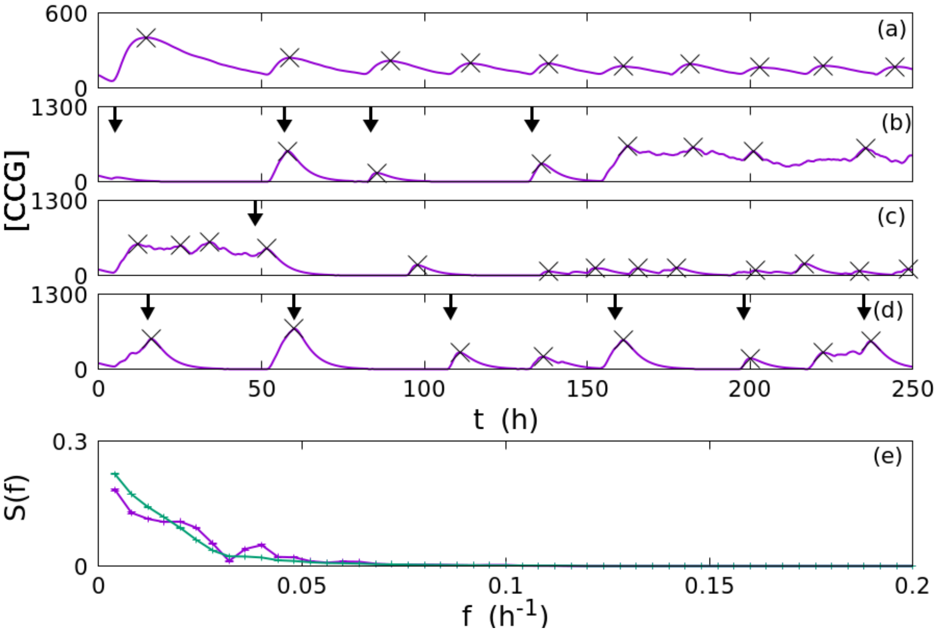

Figure 3.4: Simulation results of multi-cell hybrid model without light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[CCG]_{\text{PA,t}}$. (b)-(d) $[CCG]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[CCG]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[CCG]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

Figures 3.6-3.7 show comparisons of $[WCC]_{\text{PA,t}}$ and power spectra between different cell population sizes. It is obvious to see that the concentration PA, $[WCC]_{\text{PA,t}}$, of a cell population with smaller N exhibits more irregular oscillations than those of cell populations with larger N. This result is confirmed by results of power spectra shown in Figure 3.7.

In Figures 3.8-3.9, we used three different values of the extra-cellular coupling parameter Q for simulations of the multi-cell hybrid model without light input. The PA signals in Figure 3.8(a)-(b) show coherent oscillations but with different periods. However, the PA

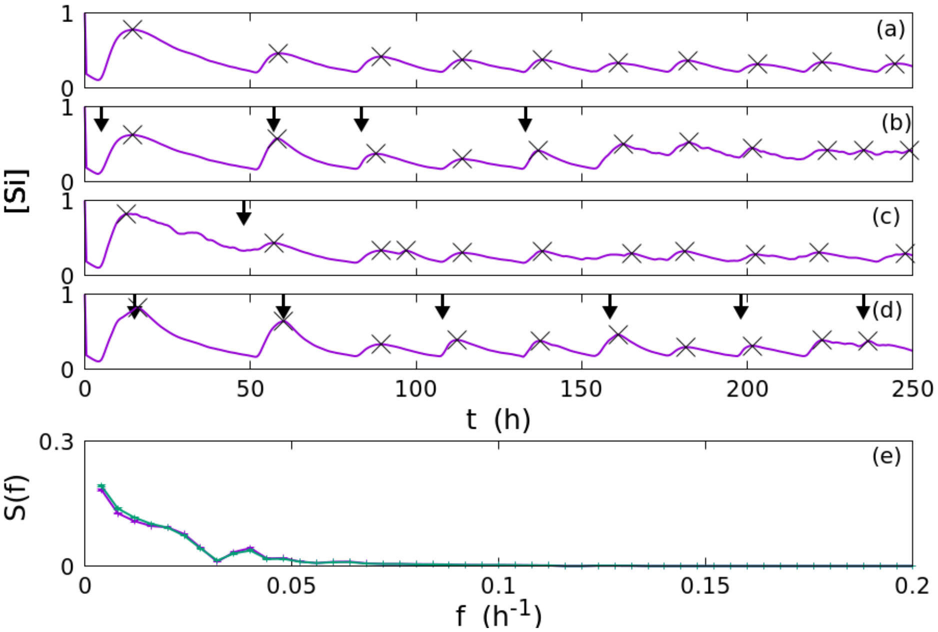

Figure 3.5: Simulation results of multi-cell hybrid model without light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[Si]_{\mathrm{PA,t}}$. (b)-(d) $[Si]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[Si]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[Si]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

signal in Figure 3.8(c), which has $Q = 0.8$, shows noisy fluctuations after cells becomes totally random. The result is consistent with the power spectra results in Figure 3.9. As we decrease Q value, sharp peaks in the power spectra of PA signals start to move to the right, indicating the decrease of period width, and finally become several broad frequency continua. This contrast of different Q values indicates that there might be a critical coupling strength, Q_{crit}, within the range of [0.8, 1.0]. If $Q \geq Q_{\text{crit}}$, cells in the population are strongly coupled

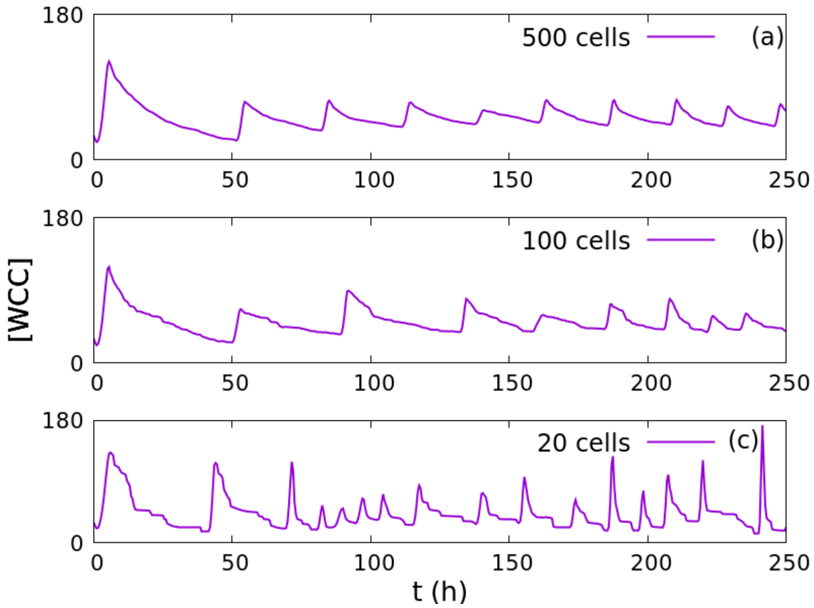

Figure 3.6: $[WCC]_{\mathrm{PA,t}}$ of different cell population sizes. Each WCC signal in (a)-(c) is obtained from only one representative, randomly chosen system trajectory.

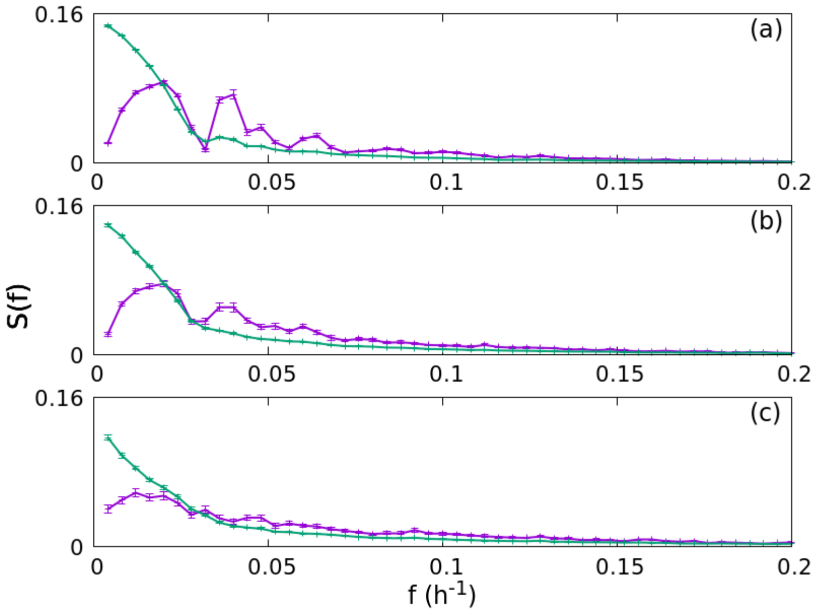

Figure 3.7: Comparisons between power spectra of the concentration PA, $[WCC]_{PA,t}$ (purple line), with PA of the power spectra of all single-cell concentrations, $[WCC]_{n,t}$ (green line). All power spectra are obtained from signals with ($t_{ini} = 0h, t_{fin} = 249.5h, J = 500$). Means and ± 1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot. (a) $N = 500$. (b) $N = 100$. (c) $N = 20$.

and build coherent and periodic PA signals. Otherwise, PA signals will become totally irregular.

Figures 3.10-3.11 show very similar results to Figures 3.8-3.9. We may assume that there is a critical coupling strength, $C_{4,\mathrm{crit}}$, above which PA signals can show coherent and periodic oscillations, within the range of [0.3,0.6].

As shown in the comparisons between single-cell and PA signals, the collective behavior in the hybrid model is quite different from the conventional synchronization process, like the Kuramoto Model, in which most of coupled oscillators share the same frequency and phase when they are in sync. From Figures 3.6-3.11, we can conclude that the cell population size, Q, and C_4 are all essential factors to the coupling strength of cells in the hybrid model.

3.3 Multi-cell system with light-entrainment

To include light in the multi-cell system, we did the same modifications to the kinetic rate equations of $[WC\text{-}1]_{n,t}$ and $[WCC]_{n,t}$ as the single-cell hybrid model by introducing the species, $Phot$, and the reaction, C_3, into the formation process of WCC. Eqs. 2.25 and 2.26 are modified for every cell, n, in a population:

$$\frac{d[WC\text{-}1]_{n,t}}{dt} = L_1 \cdot [wc\text{-}1^{r1}]_{n,t} - (D_4 + C_2) \cdot [WC\text{-}1]_{n,t} - Phot \cdot C_3 \cdot [WC\text{-}1]_{n,t} \tag{3.19}$$

$$\begin{aligned}\frac{d[WCC]_{n,t}}{dt} &= C_2 \cdot [WC\text{-}1]_{n,t} - (D_8 + P \cdot [FRQ]_{n,t}{}^{m_{\mathrm{F}}}) \cdot [WCC]_{n,t} \\ &\quad + Phot \cdot C_3 \cdot [WC\text{-}1]_{n,t}\end{aligned} \tag{3.20}$$

Figures 3.12-3.2 display some of the simulation results of the multi-cell hybrid model with a 12+12h light input, which is a 24h artificial day. Each Panel (a) shows the concentration PA, $[X]_{\mathrm{PA,t}}$, of a random system trajectory of a $N = 500$-cell population. Panels (b)-(d) show three randomly selected intra-cell concentrations, $[X]_{n,t}$, which are from the corresponding

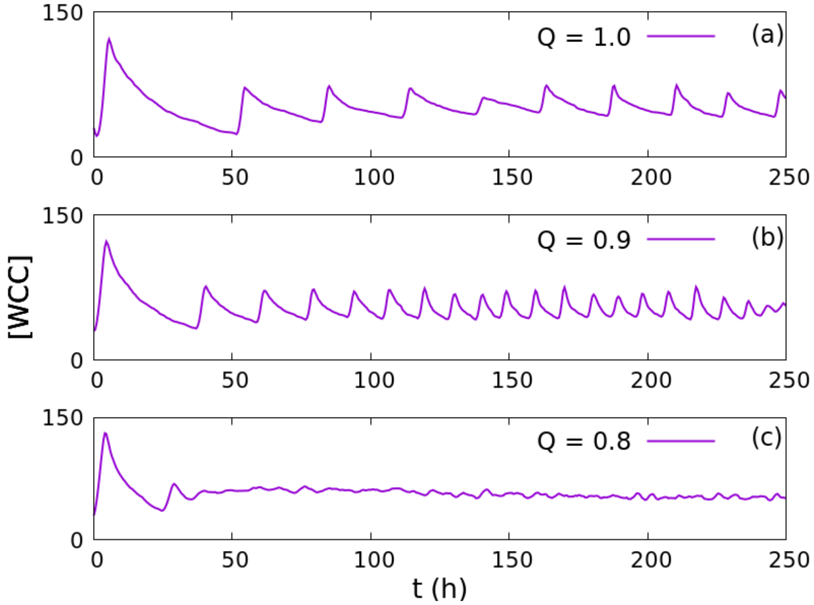

Figure 3.8: $[WCC]_{\mathrm{PA,t}}$ of different values of coupling parameter Q. Each WCC signal in (a)-(c) is obtained from only one representative, randomly chosen system trajectory.

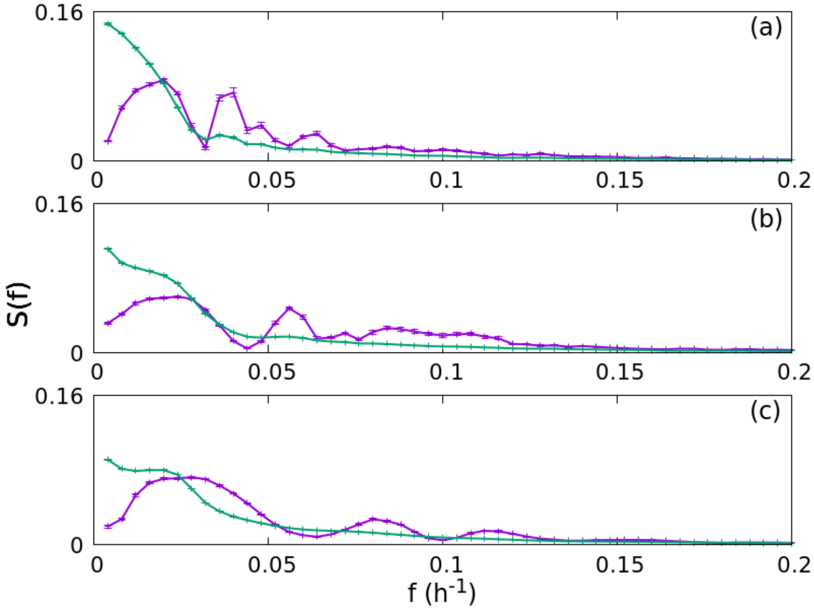

Figure 3.9: Comparisons between power spectra of the concentration PA, $[WCC]_{PA,t}$ (purple line), with PA of the power spectra of all single-cell concentrations, $[WCC]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot. (a) $Q = 1.0$. (b) $Q = 0.9$. (c) $Q = 0.8$.

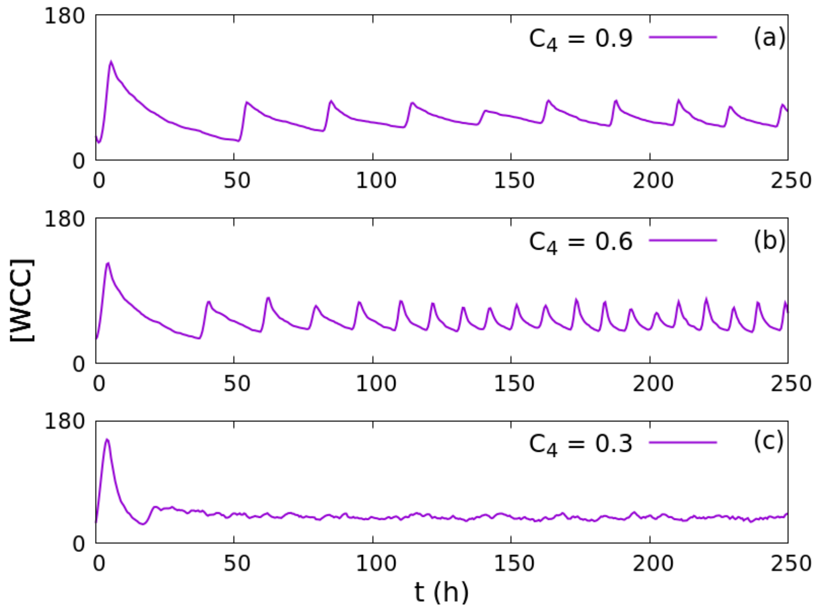

Figure 3.10: $[WCC]_{\mathrm{PA,t}}$ of different values of coupling parameter C_4. Each WCC signal in (a)-(c) is obtained from only one representative, randomly chosen system trajectory.

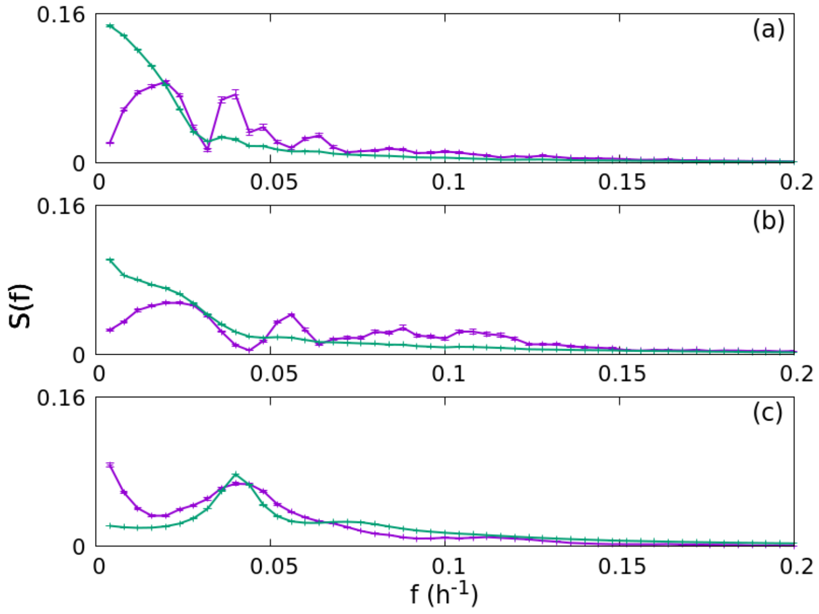

Figure 3.11: Comparisons between power spectra of the concentration PA, $[WCC]_{PA,t}$ (purple line), with PA of the power spectra of all single-cell concentrations, $[WCC]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot. (a) $C_4 = 0.9$. (b) $C_4 = 0.6$. (c) $C_4 = 0.3$.

populations of Panel (a). The ×-symbols in Panels (a)-(d) represent the local maxima in the curves. The vertical arrows in Panels (b)-(d) indicate the short frq-gene bursts. Error bars in each Panel (e) are ±1 SEM, generated by the Bootstrap sampling method for a random sample of $N_{\mathrm{R}} = 100$ system trajectories. The phasing of the light exposure for all simulations in the section are L/D.

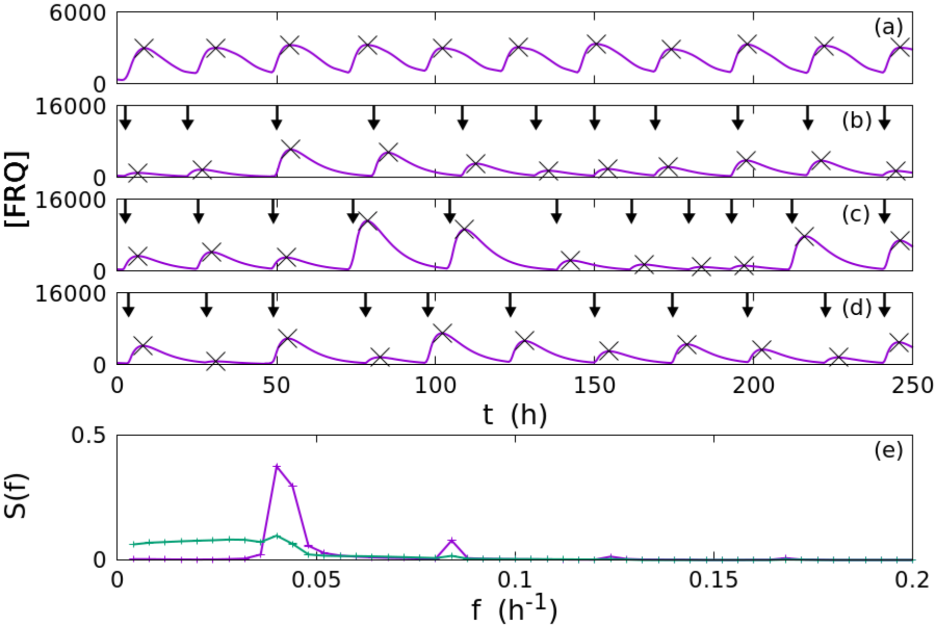

Figure 3.12: Simulation results of multi-cell hybrid model with a 12+12h light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[FRQ]_{\mathrm{PA,t}}$. (b)-(d) $[FRQ]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[FRQ]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[FRQ]_{n,t}$ (green line). All power spectra are obtained from signals with ($t_{ini} = 0h, t_{fin} = 249.5h, J = 500$). Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

In Figures 3.12-3.15, it is obvious to see that most of cells have shown clear responses to the 12+12h light input. The number of maxima for each intra-cell concentration, $[X]_{n,t}$, is nearly the same as the number of maxima in the PA concentration,$[X]_{\mathrm{PA,t}}$. In spite of their incoherent amplitudes, single-cell signals show approximately periodic oscillations,

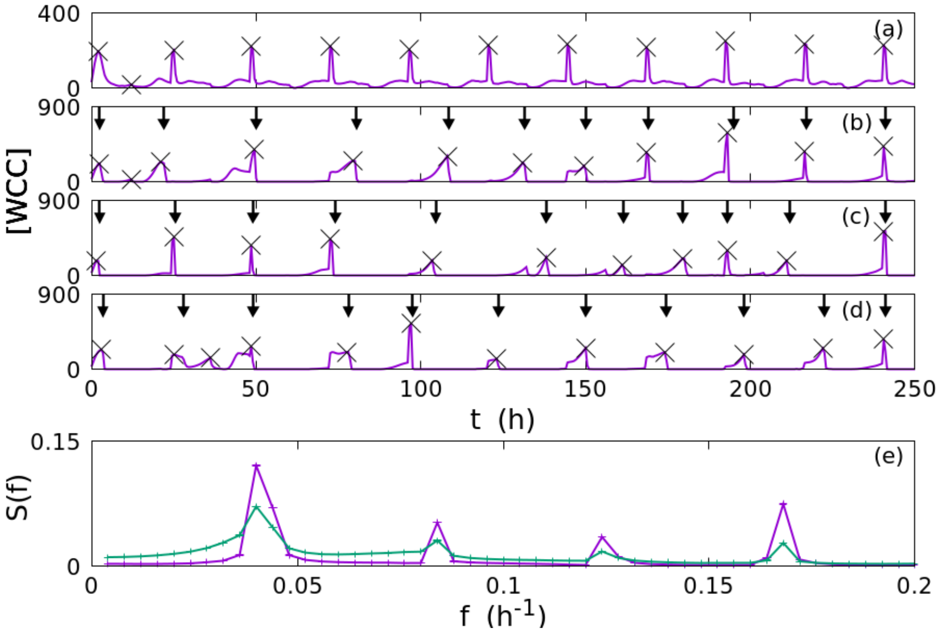

Figure 3.13: Simulation results of multi-cell hybrid model with a 12+12h light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[WCC]_{\mathrm{PA,t}}$. (b)-(d) $[WCC]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[WCC]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[WCC]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

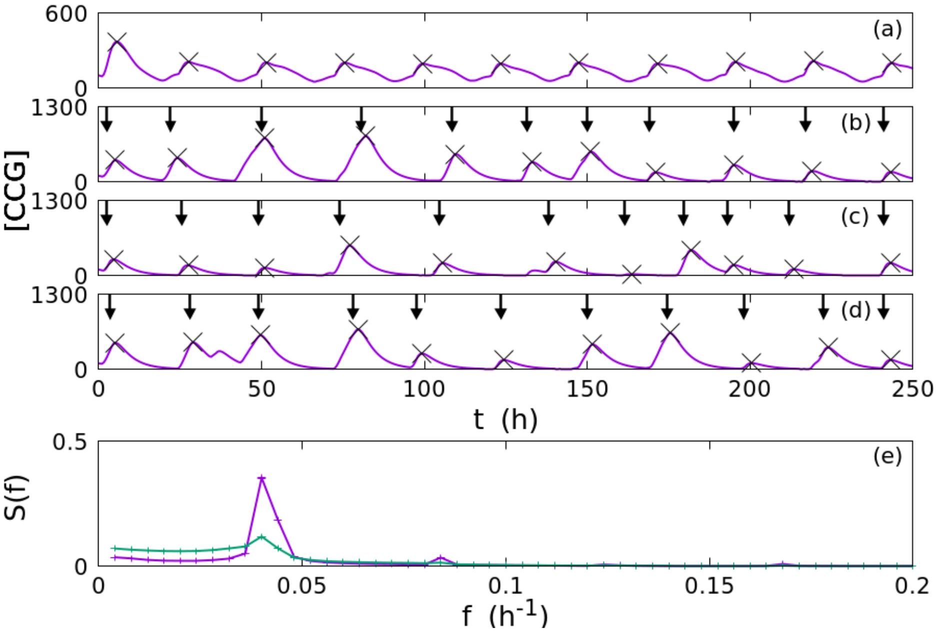

Figure 3.14: Simulation results of multi-cell hybrid model with a 12+12h light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[CCG]_{\mathrm{PA,t}}$. (b)-(d) $[CCG]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[CCG]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[CCG]_{n,t}$ (green line). All power spectra are obtained from signals with ($t_{ini} = 0h, t_{fin} = 249.5h, J = 500$). Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

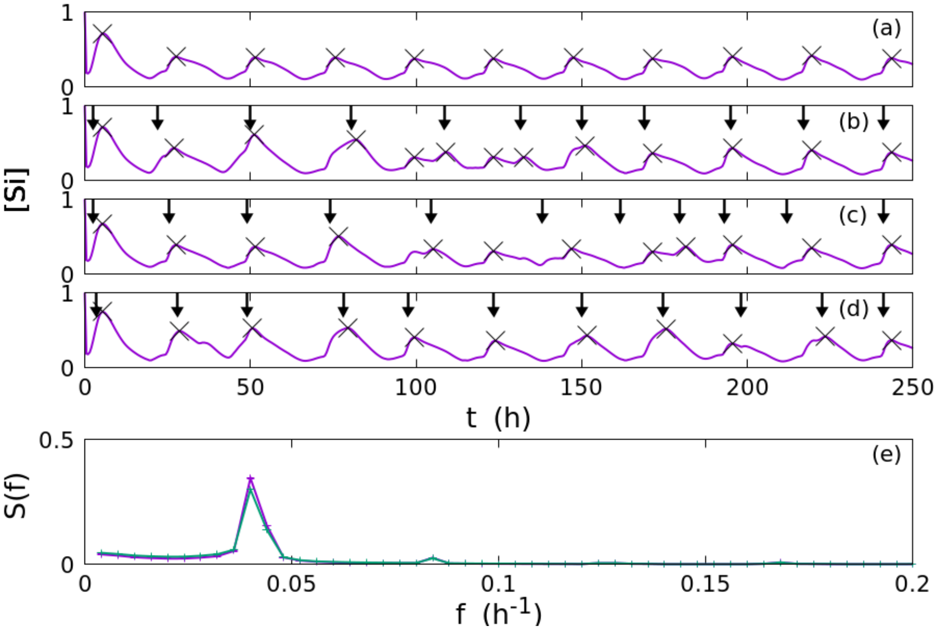

Figure 3.15: Simulation results of multi-cell hybrid model with a 12+12h light input. All signals in (a)-(d) are obtained from only one representative, randomly chosen system trajectory. The ×-symbols in panels (a)-(d) mark the dominant local maxima, defined in section 4.1. (a) Concentration PA, $[Si]_{\mathrm{PA,t}}$. (b)-(d) $[Si]_{n,t}$ of three randomly chosen single cells, n. (e) Power spectrum of the concentration PA, $[Si]_{PA,t}$ (purple line), and PA of the power spectra of all single-cell concentrations, $[Si]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars of $S(f)$ may be smaller than the plotted line width and therefore not visible in the plot.

which are almost in sync with PA signals. This synchronization is also shown in the power spectra results in Panels (e) of Figures 3.12-3.15, as both purple and green curves have their fundamental frequencies at around $f = 0.417h^{-1}$, which is exactly the frequency of the light pulse.

In Figures 3.16-3.17, simulations with three different light inputs, 6+6h, 12+12h, and 18+18h were performed. The model shows very clear responses to different lengths of light inputs, for both PA and single-cell signals. Figure 3.17 presents a quite different results comparing with results shown in Figure 2.6, where we may assume that the coupling mechanism enhances the system's responsiveness to light stimuli.

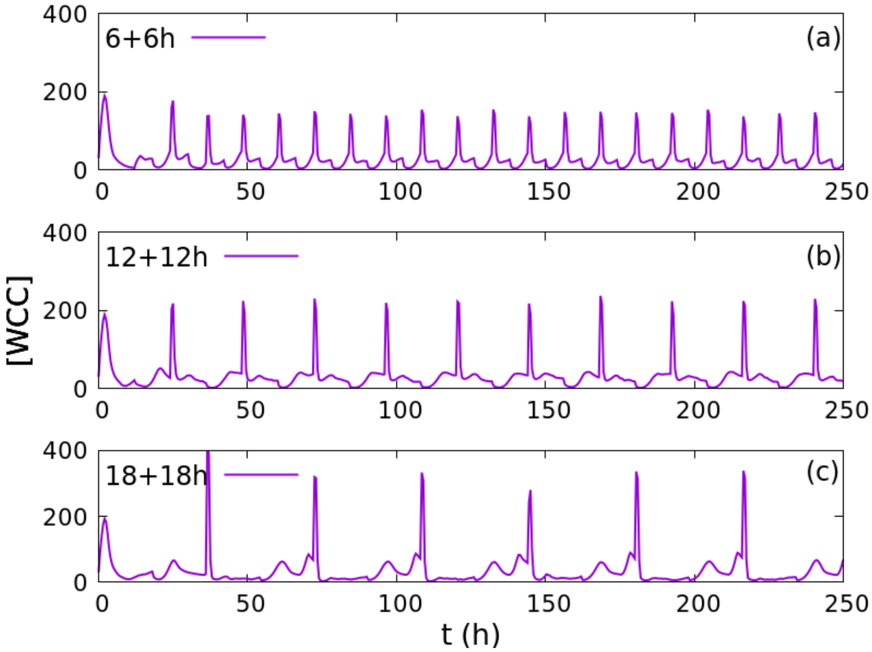

Figure 3.16: $[WCC]_{\mathrm{PA,t}}$ of different light inputs. Each WCC signal in (a)-(c) is obtained from only one representative, randomly chosen system trajectory. (a) 6+6h. (b) 12+12h. (c) 18+18h.

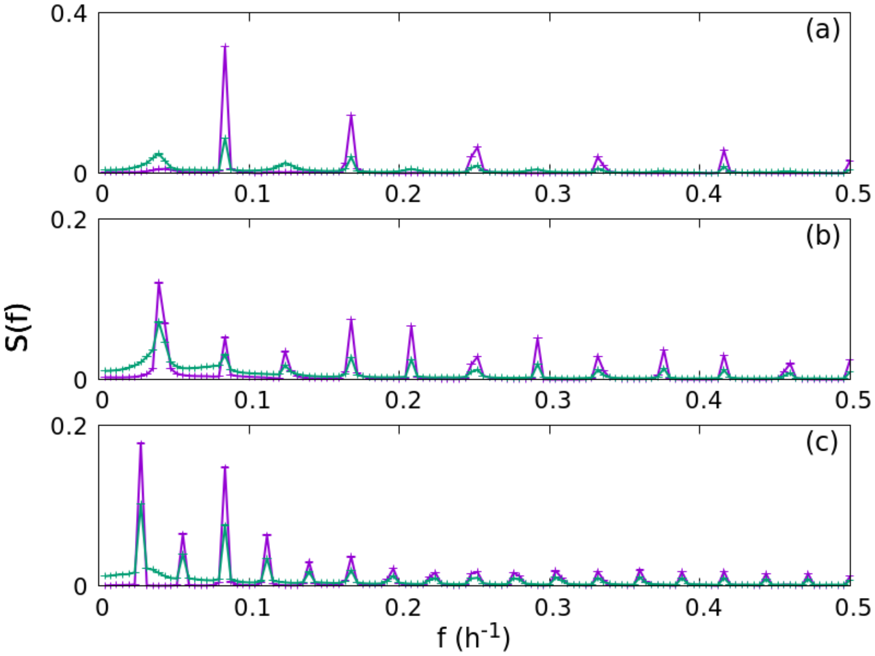

Figure 3.17: Comparisons between power spectra of the concentration PA, $[WCC]_{PA,t}$ (purple line), with PA of the power spectra of all single-cell concentrations, $[WCC]_{n,t}$ (green line). All power spectra are obtained from signals with $(t_{ini} = 0h, t_{fin} = 249.5h, J = 500)$. Means and ±1 SEM error bars are obtained by bootstrapping a sample of $N_R = 100$ system trajectories and $B = 1000$ bootstrap samples. (a) 6+6h. (b) 12+12h. (c) 18+18h. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

Chapter 4

Synchronization measures for stochastic coupled clock oscillators

4.1 Statistics for the collective behavior of coupled oscillators in hybrid model

4.1.1 Window maximization filter

As the simulation results shown in the Chapter 3, the concentration population average (PA) presents smooth and nearly periodic oscillations, while most of the single-cell oscillators in the population are actually out of phase. To study this distinctive collective behavior, we want to introduce several methods to characterize some of the statistical traits caused by the interactions between the coupled oscillators in the hybrid model.

The first approach, which is fundamental to the rest of the methods, is called *window maximization* (WinMax) filter. It, basically, uses a non-linear time-domain low-pass signal filter to completely remove all short-time fluctuations from a noisy curve, like our highly stochastic single-cell concentration time series. When applying this approach, we continuously slide a short-time window, whose size is T_{W}, from the beginning of the trajectory to its end with the same interval as the one of the concentration time series. At each step of the sliding process, a maximum value within the window's range is chosen as the new filtered signal for the time point of the window's head.

As a result, local maxima or minima, which are referred to as T_{W}-*dominant* local (DL) maxima or minima, of the original concentration time series can be accurately extracted by simply searching for plateaus or basins in the filtered output signal. In other words, suppose a local maximum of the concentration time series, $y(t)$, is occurring at some time t_m. Such a maximum will be called, by definition, a *dominant local* (DL) maximum if it satisfies the condition $y(t_m) \geq y(t)$ for all times t within a distance of T_W from t_m, i.e., for all t obeying $|t - t_M| < T_W$. DL minima are defined analogously, with the foregoing inequality replaced by $y(t_m) \leq y(t)$. These DL maxima and minima are then used to characterize how individual cells contribute - or not - to the coherent diurnal PA signal. Panels (a)-(d) in Figures. 3.2-3.5 and 3.12-3.15 above illustrate DL maxima, shown as $\times - symbols marking the time series data.$

In the foregoing figures in Chapter 3, and also for all other DL maximum or DL minimum calculations below, we have consistently used a short-time cut-off value of $T_W = 5h$ for models in the dark, which is comparable to about one quarter of the lesser max-to-max spacings seen in those figures for PA time series, and $T_W = 6h$ for models with a 12+12h light exposure. In contrast to more conventional linear filtering methods [29], the location, t_m, of the DL maxima and DL minima identified with the WinMax approach does not change when T_W is varied over some range of values, since every DL maximum, t_m, is a local maximum of y(t): as long as WinMax identifies t_m as a DL maximum, the value of t_m does not vary when T_W is varied. For example, over some range, say, from $T_W = 4h$ to $T_W = 8h$, in the case of the hybrid model discussed here, the WinMax-based results hardly change at all with T_W. In general, the locations, t_m, identified as DL maxima or DL minima, do not vary significantly, as long as T_W is less than about one half of the lesser DL max-to-max or, respectively, DL min-to-min spacings, and larger the short time scales of the rapid random fluctuations to be suppressed by the WinMax. The any variations that occur with changing T_W are either t_m-values that are disqualified as DL extrema, if T_W is made too large, or else, t_m-values that are admitted as additional DL extrema, when T_W is made too small.

4.1.2 Clock-readout

In Figure 3.2-3.5, most single-cell signals without any input light show that the timing of their DL maxima is quite random and irregular, but PA signals, in contrast, have coherently timed DL maxima. To explicitly analyze the difference between the clock running speeds reflected in the two different signals, we introduced a curve, referred to as the *clock-readout*, t_m, to exhibit the timings of consecutive DL maxima as a function of the ascending maxima number, m.

Figure 4.1 provides the comparisons of the clock-readouts of the multi-cell hybrid model with dark between single-cell signals and PA signals of $[FRQ]$, $[WCC]$, $[CCG]$, and $[Si]$. The single-cell DL maxima of $[FRQ]$ and $[WCC]$, on average, are spaced about twice further apart than their PA maxima, which indicates that the single-cell clocks of $[FRQ]$ and $[WCC]$ run much slower than the PA diurnal clocks of $[FRQ]$ and $[WCC]$. By contrast, the single-cell clocks of $[CCG]$ and $[Si]$ run almost as fast as their PA diurnal clocks. However, the standard deviations (SDs) of all the 4 species' single-cell clock-readouts are much larger than those of PA clock-readouts, which strongly proves that the oscillations of single-cell signals are much more irregular than PA signals' oscillations. The presence of linearity and small SD values in the PA clock-readout curves show that the cell population as a whole is a more stable and reliable clock than the individual cells within the population. The slopes and intercepts of linear fitted lines [30] to the PA clock-readout data in Figure 4.1 from $m = 1$ to $m = 10$ are listed in Table 4.1. Slopes in Table 4.1 provide estimates of the clock oscillation periods, which are consistent with experimentally observed periods, around 23h [7, 21–23], seen in dark-running N. crassa clocks.

The clock-readouts of the multi-cell hybrid model with a light input of 12+12h artificial day is shown in Figure 4.2. The phasing of the light exposure for all simulations in the section are L/D. Due to the influence of light input, the individual cell DL maxima of all the four species are spaced quite close to the PA maxima. In addition, SDs of both the individual cell and PA clock-readouts of the four species are considerably reduced, comparing with those in Figure 4.1. A nearly constant max-to-max spacing is exhibited in the PA curve of each species, along with the straight fitted line, which is right on the top of the PA curve of each species. The slopes of the straight fitted lines are shown in Table 4.2, which are almost the same as the period of the artificial day.

By bootstrapping the base samples of $N_R = 100$ random trajectories into $B = 1000$ bootstrap samples, we find that the SEM values of the mean t_m results for both single-cell signals and PA signals are less than 5h and 1h for Figures 4.1 and 4.2, respectively.

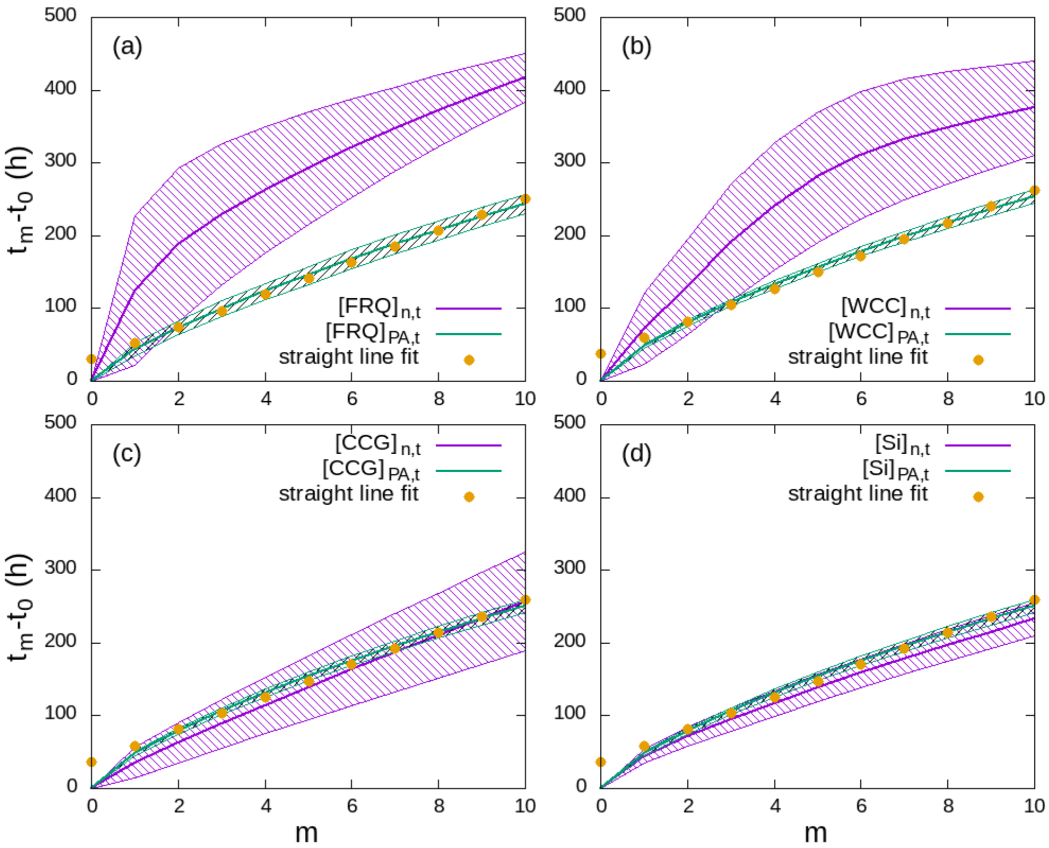

Figure 4.1: The clock-readout, t_m, of DL maxima is plotted vs the index, m, of DL maxima for 4 species of the multi-cell hybrid model in the dark. All DL maxima of both single-cell signals and PA signals are extracted by the WinMax filter with $T_{\mathrm{W}} = 5h$. The purple and green lines are the means of single-cell and PA t_m, respectively. Shading is ±1 SD. Means and SDs of t_m are calculated by from a random sample of $N_{\mathrm{R}} = 100$ system trajectories of a $N = 500$-cell population, generated by using the default uniform initial conditions and reaction rate parameter set from Table 2.2, 2.1, and 3.1. The yellow dots indicates the linear fitted lines for the means of PA signals' DL maxima.

Table 4.1: Slopes and intercepts of fitted curves to the means of PA clock-readouts for the multi-cell hybrid model in the dark. Only the t_m-data points from $m = 1$ to $m = 10$ in Figure. 4.1 were included in the straight line fit. Means and SEMs are obtained from bootstrapping $B = 1000$ bootstrap samples from a base sample of $N_R = 100$ system trajectories.

	FRQ		WCC		CCG		Si	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Slope	21.96	0.15	22.57	0.11	22.37	0.10	22.37	0.11
Intercept	29.86	0.98	36.28	0.37	35.28	0.36	35.27	0.37

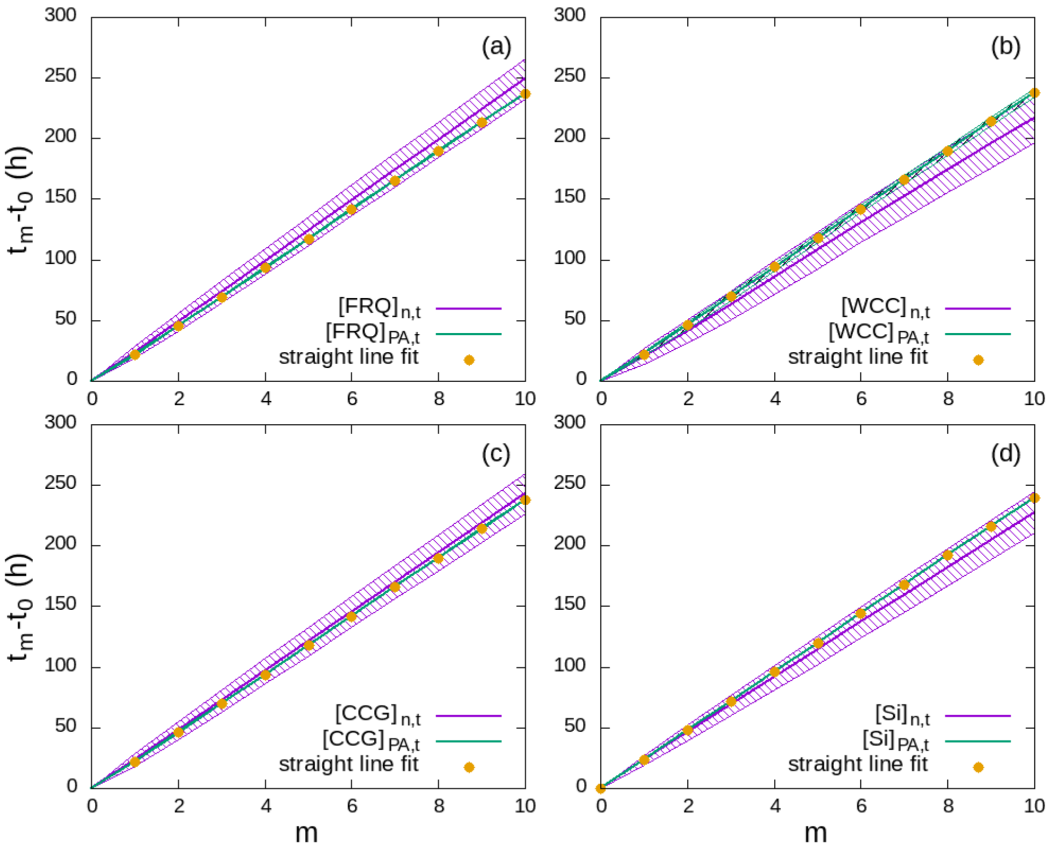

Figure 4.2: The clock-readout, t_m, of DL maxima is plotted vs the index, m, of DL maxima for 4 species of the multi-cell hybrid model with a 12+12h light exposure. All DL maxima of both single-cell signals and PA signals are extracted by the WinMax filter with $T_W = 6h$. The purple and green lines are the means of single-cell and PA t_m, respectively. Shading is ±1 SD. Means and SDs of t_m are calculated from a random sample of $N_R = 100$ system trajectories of a $N = 500$-cell population, generated by using the default uniform initial conditions and reaction rate parameter set from Table 2.2, 2.1, and 3.1. The yellow dots indicates the linear fitted lines for the means of PA signals' DL maxima.

Table 4.2: Slopes and intercepts of fitted curves to the means of PA clock-readouts for the multi-cell hybrid model with a 12+12h light exposure. Only the t_m-data points from $m = 1$ to $m = 10$ in Figure 4.2 were included in the straight line fit. Means and SEMs are obtained from bootstrapping $B = 1000$ bootstrap samples from a base sample of $N_R = 100$ system trajectories.

	FRQ		WCC		CCG		Si	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Slope	23.985	0.003	23.95	0.01	24.000	0.001	24.001	0.001
Intercept	-2.54	0.03	-1.65	0.27	-2.04	0.01	-0.001	0.003

4.1.3 Beat-skip probability

To study the relation between an individual cell' participation in two successive DL maxima of the PA signal of a molecular species, X, we tried to measure the probability, named "beat-skip" probability, for a cell n to skip the $(m+j)$-th DL maximum at $t_{m+j,\mathrm{PA}}$ in $[X]_{\mathrm{PA,t}}$, given that it has a contribution to the m-th DL maximum at $t_{m,\mathrm{PA}}$ in$[X]_{\mathrm{PA,t}}$. With the help of the WinMax filter introduced above, all maxima in both single-cell and PA signals can be easily located. To determine whether a cell n has contributed to the DL maximum at $t_{m,\mathrm{PA}}$ in $[X]_{\mathrm{PA,t}}$, we used a participation time range, T_{P}, to see if the cell n has a maximum in $[X]_{n,t}$ at $t_{m',n}$, which obeys $|t_{m',n} - t_{m,\mathrm{PA}}| \leq T_{\mathrm{P}}$. If so, the cell is counted as one contribution to the m-th DL maximum in $[X]_{\mathrm{PA,t}}$. The 2-dimensional beat-skip probability, which is conditional, is shown below:

$$p_{\mathrm{BS}}^{(2D)}(m+j|m) = \frac{N(m+j|m)}{N(m)}, \tag{4.1}$$

where $N(m+j, m)$ is the number of cells which contribute to the m-th DL maximum of the PA signal but not to the $(m+j)$-th DL maximum, and $N(m)$ is the number of cells which contribute to the m-th DL maximum of the PA signal.

Equation 4.1 can be condensed into a one dimensional form,

$$p_{\mathrm{BS}}(j) = \frac{1}{n_{max} - j} \sum_{m=1}^{n_{max}-j} p_{\mathrm{BS}}^{(2D)}(m+j, m), \tag{4.2}$$

where n_{max} is the total number of DL maxima in the PA signal.

Figure 4.3 displays the means and error bars for the beat-skip probabilities of species *FRQ*, *WCC*, *CCG*, and *Si* of the multi-cell hybrid model with dark. Here, we choose $T_{\mathrm{P}} = 5h$, which is almost the diurnal quarter period. Most of the species, except for *Si*, have shown large values of $p_{BS}(j)$, which means that an individual cell always has large probabilities, mostly more than 40%, to skip the following PA maxima, when it contributes to a certain PA maximum in the first place. This is consistent with the result of clock-readouts that individual cells' clock runs much slower than the PA maxima, since the PA signal always has more number of maxima than single-cell signals it contains.

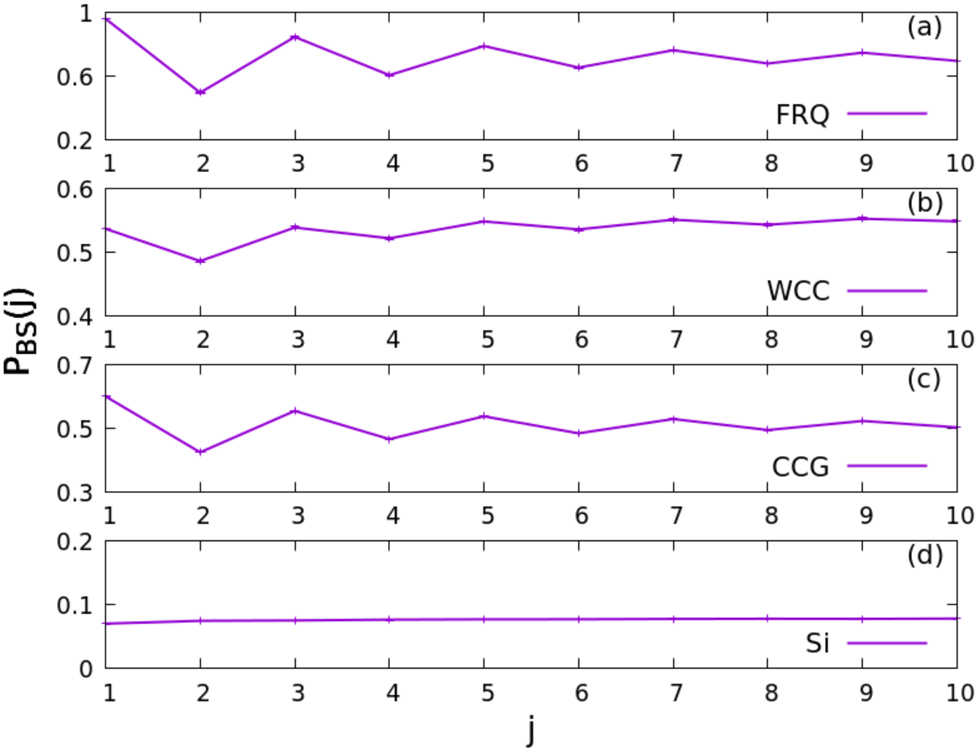

Figure 4.3: The beat-skip probability, $p_{BS}(j)$, is plotted vs the time lag index j for 4 species of the multi-cell hybrid model in the dark. Means and ±1 SEM error bars of $P_{BS}(j)$ are calculated by bootstrapping a random sample of $N_{\mathrm{R}} = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

When we apply the light input to the multi-cell hybrid model, the beat-skip probability results are displayed in Figure 4.4. The phasing of the light exposure for all simulations in the section are L/D. The participation time range is a quarter of the artificial day period, i.e. $T_{\mathrm{P}} = 6h$. Because of the influence of the input light, the values of $p_{\mathrm{BS}}(j)$ for FRQ, WCC, and CCG are considerably reduced, which means that an individual cell has much lower probabilities to skip subsequent PA maxima after it contributes one, when comparing with the dark hybrid model.

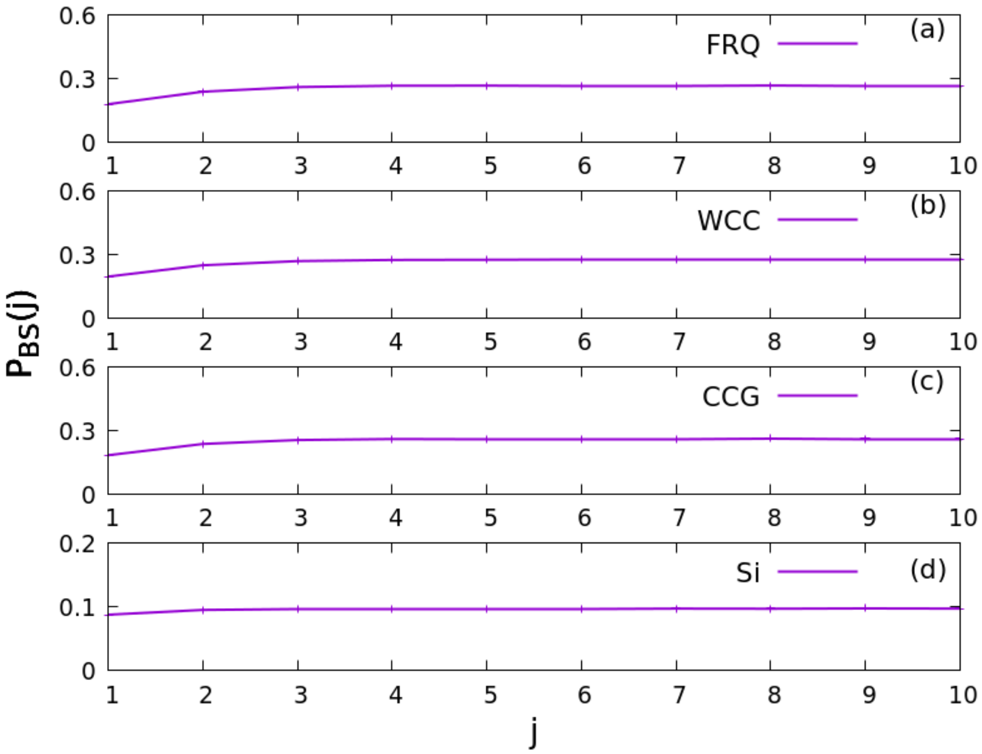

Figure 4.4: The beat-skip probability, $p_{BS}(j)$, is plotted vs the time lag index j for 4 species of the multi-cell hybrid model with a 12+12h light exposure. All DL maxima of both single-cell signals and PA signals were extracted by the WinMax filter with $T_{\mathrm{W}} = 6h$. Means and ± 1 SEM error bars of $p_{BS}(j)$ are calculated by Bootstrapping a random sample of $N_{\mathrm{R}} = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

4.1.4 Pearson correlation

Another way to analyze the relation between an individual cell's participation to two successive PA maxima is to measure the Pearson correlation of them. First, we define a participation score, $s_{m,n}$, for a cell n relative to the m-th DL maximum, at $t_{m,\mathrm{PA}}$, in the PA signal of a molecular species, $X_{\mathrm{PA,t}}$. If the cell n has a maximum occurs at the time that is within the participation range T_{P}, as described in the definition of the beat-skip probability, from $t_{m,\mathrm{PA}}$, we can set $s_{m,n} = 1$. Otherwise we set $s_{m,n} = 0$. With the definition of the participation scores, we can calculate the Pearson correlation [31] of $s_{m,n}$ and $s_{m+j,n}$:

$$C_{\mathrm{PC}}^{(2D)}(m, m+j) = \frac{\sum_{n=1}^{N} \Delta s_{m,n} \cdot \Delta s_{m+j,n}}{\sqrt{\left(\sum_{n=1}^{N} \left(\Delta s_{m,n}\right)^2 \cdot \sum_{n'=1}^{N} \left(\Delta s_{m+j,n'}\right)^2\right.}}, \tag{4.3}$$

where $\Delta s_{m,n} = s_{m,n} - \frac{1}{N} \sum_{n'=1}^{N} s_{m,n'}$, and N is the total number of cells in the population. The 1-dimensional Pearson correlation, as a function of j, is presented below:

$$C_{\mathrm{PC}}(j) = \frac{1}{n_{max} - j} \sum_{m=1}^{n_{max}-j} C_{\mathrm{PC}}^{(2D)}(m+j, m), \tag{4.4}$$

where n_{max} is the total number of DL maxima in the PA signal.

Figure 4.5 shows the means, denoted by $p_{\mathrm{P}}(m)$, and error bars of participation scores, with $T_{\mathrm{P}} = 5h$, for species *FRQ*, *WCC*, *CCG*, and *Si* of the multi-cell hybrid model in the dark environment. It is easily to see that the participation rates almost stay at very small values for *FRQ* ($\leq 20\%$), *WCC* ($\leq 50\%$), and *CCG* ($\leq 50\%$), with $m \geq 2$ or 4, despite the fact that all cells started with the same initial conditions at the beginning.

Figure 4.6 shows the corresponding means and error bars of Pearson correlation $C_{\mathrm{PC}}(j)$ to Figure 4.5. The small values of $C_{\mathrm{PC}}(j)$ for *FRQ*, *WCC*, and *CCG* indicate that the contributing sub-populations for the PA maxima are quite unstable. In other words, the cells which contribute to the DL maxima of a PA signal keep changing from the start of the oscillation to its end. The results of $p_{\mathrm{P}}(m)$ and C_{PC}, along with $p_{\mathrm{BS}}(j)$, are consistent with the what are shown in the plots of clock-readouts.

Figures 4.7 and 4.8 are results of $p_{\mathrm{P}}(m)$ and $C_{\mathrm{PC}}(j)$, respectively, by using the multi-cell hybrid model with a 12+12h light input. The phasing of the light exposure for all simulations in the section are L/D. The participation period is $T_{\mathrm{P}} = 6h$. Due to the regular light input, the participation rates of all species are around or more than 80%. In addition, the Pearson correlation of $s_{m,n}$ and $s_{m+j,n}$ goes to zero with $m \geq 3$, which indicates that the large values

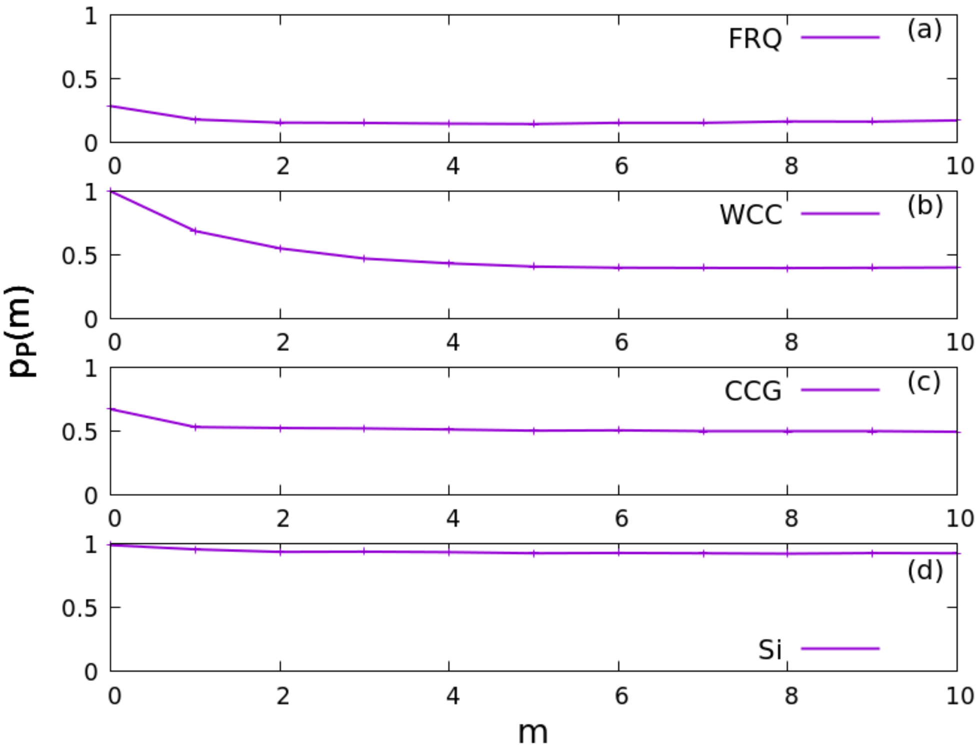

Figure 4.5: Mean participation score, $p_{\mathrm{P}}(m)$, vs. DL maximum index, m, of concentration PA for 4 species of the multi-cell hybrid model in the dark. Means and ±1 SEM error bars of $p_{BS}(j)$ are calculated by bootstrapping a random sample of $N_{\mathrm{R}} = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

of contributing sub-populations are mostly driven by the light input, and they have little to do with the PA maxima in the vicinity.

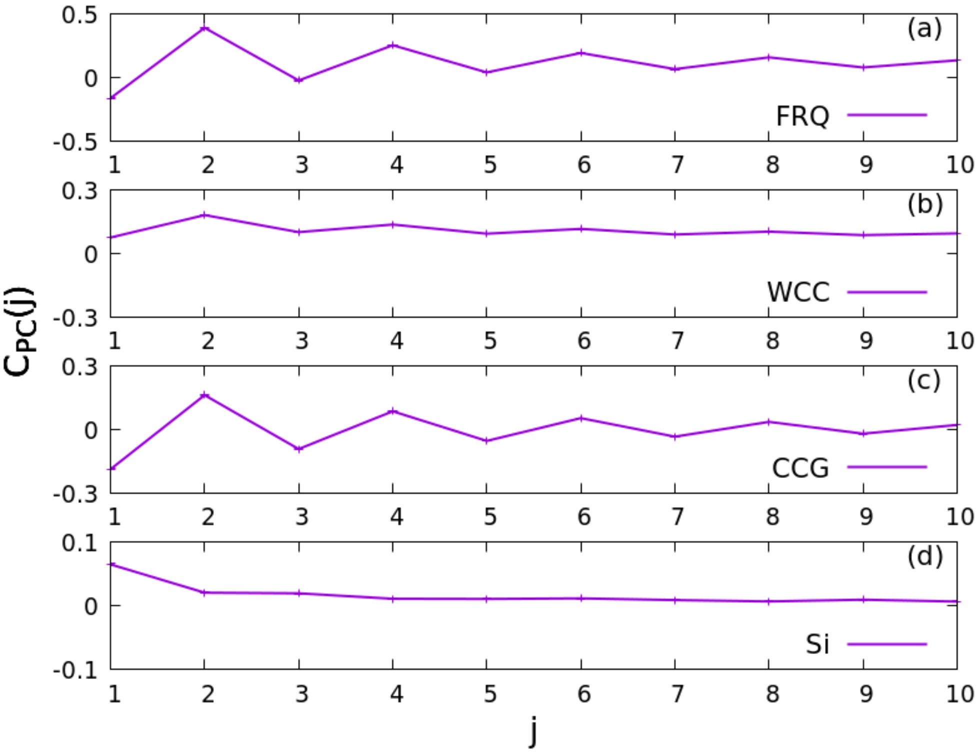

Figure 4.6: Pearson correlation, C_{PC}, vs. the time lag index j for 4 species of the multi-cell hybrid model in the dark. Means and ± 1 SEM error bars of $C_{\mathrm{PC}}(j)$ are calculated by bootstrapping a random sample of $N_{\mathrm{R}} = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

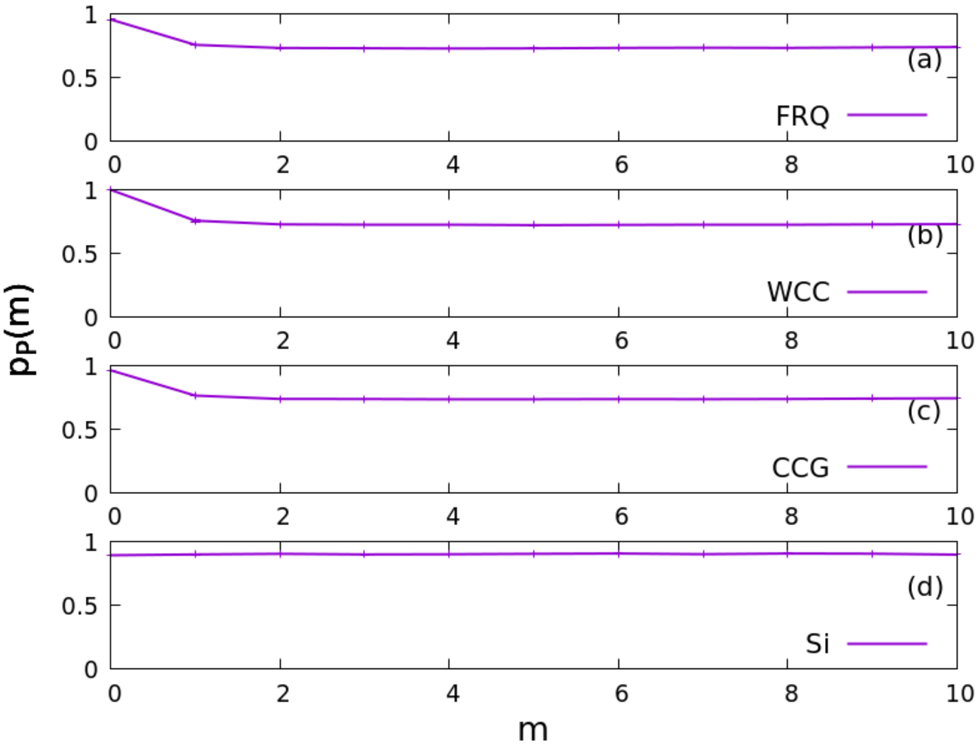

Figure 4.7: Mean participation score, $p_{\mathrm{P}}(m)$, vs. DL maximum index, m, of concentration PA for 4 species of the multi-cell hybrid model with a 12+12h light exposure. Means and ± 1 SEM error bars of $p_{BS}(j)$ are calculated by bootstrapping a random sample of $N_{\mathrm{R}} = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

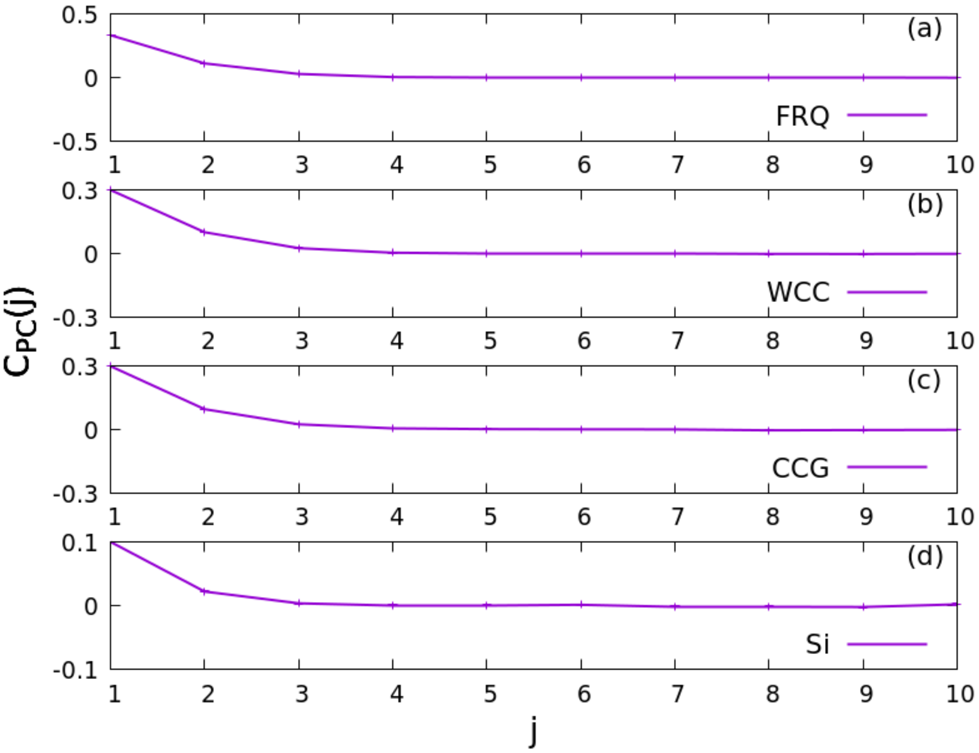

Figure 4.8: Pearson correlation, C_{PC}, vs. the time lag index j for 4 species of the multi-cell hybrid model with a 12+12h light exposure. Means and ±1 SEM error bars of $C_{\mathrm{PC}}(j)$ are calculated by bootstrapping a random sample of $N_{\mathrm{R}} = 100$ system trajectories and $B = 1000$ bootstrap samples. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

4.2 Definition of the synchronization order parameter

The analyses in the preceding section indicate that the cell population as a whole utilizes the single-cell stochasticity as a foundation for self-organization of coherent PA oscillations, instead of working against it. Moreover, the clock system presents a high level of robustness for its oscillatory behavior at a substantial coupling strength, where $Q = 1.0$ and $C_4 = 0.9$. To quantify this oscillatory behavior and see if it can still be observed with different coupling parameters, we introduce a synchronization order parameter, F_X, to measure the amplitude of the PA oscillations relative to the overall strength of the PA signal, for any species X. Unlike traditional synchronization measures, e.g. Kuramoto model order parameter [32] and phase locking value [33], which mainly uses the phase-coherence of oscillators in the population to represent the extent of synchronization, we only focus on the PA signal in the calculation of F_X. To clearly define F_X, the whole process is split into several steps, as shown below:

First, the WinMax filter is used to locate all DL maxima at t^+_m and DL minima at t^-_m,immediately following t^+_m, within a finite time interval $[0, T_{obs}]$, where T_{obs} is 500h in our simulation.

Second, we want to find maxima, m_{ini} and m_{fin}, which are the first maxima in the time ranges of $[T_{ini}, T_{fin})$ and $[T_{fin}, T_{obs}]$, respectively, by using

$$m_{ini} := \begin{cases} min(m|T_{ini} \leq t^+_m < T_{fin}) & (m|T_{ini} \leq t^+_m < T_{fin}) \text{ is not empty} \\ -1 & \text{otherwise} \end{cases} \tag{4.5}$$

$$m_{fin} := \begin{cases} min(m|T_{fin} \leq t^+_m \leq T_{obs}) & (m|T_{fin} \leq t^+_m \leq T_{obs}) \text{ is not empty} \\ -1 & \text{otherwise} \end{cases} \tag{4.6}$$

where $T_{ini} = 125h$ and $T_{fin} = 250h$ in the simulation. Based on the locations of maxima m_{ini} and m_{fin}, we can define $t_{ini} = t^+_{m_{ini}}$ and $t_{fin} = t^+_{m_{fin}}$. In this step, a signal segment within the time range of $[t_{ini}, t_{fin}]$, which has $M_T + 1$ maxima, is extracted from PA signal, $[X]_{\mathrm{PA,t}}$, for later analysis.

Third, a variable transformation from time, t, to pseudo-phase variable, ϕ, is introduced for $t \in [t_{ini}, t_{fin}]$, by

$$\phi(t) \;:=\; \begin{cases} \pi \cdot (2m - 2m_{ini}) + \pi \cdot \dfrac{t - t^+_m}{t^-_m - t^+_m} & \text{if } t \in [t^+_m, t^-_m] \\ \pi \cdot (2m - 2m_{ini} + 1) + \pi \cdot \dfrac{t - t^-_m}{t^+_{m+1} - t^-_m} & \text{if } t \in [t^-_m, t^+_{m+1}] \end{cases} \tag{4.7}$$

where $m = m_{ini}, m_{ini} + 1, ..., m_{fin} - 1$. The variable transformation of this step sets the segment signal's max-to-max distance to a constant value, 2π, in ϕ space.

Fourth, by using the inverse transformation $t(\phi)$, which can be derived from Equation 4.7, we can define the ϕ-transformed concentration PA, $[X]_{\mathrm{PA},\phi} := Z(\phi)$, in the ϕ-space. Because the signal data are discretely collected in time space, an equidistant ϕ-grid with $N_\phi = M_T \cdot M_\phi$ grid intervals can be introduced to evaluate discrete signal data, $Z(\phi_j)$, in the ϕ space by means of linear interpolation from underlying equidistant t-grid to the non-equidistant grid $t(\phi_j)$. In the simulation, the number of grid intervals between two adjacent maxima is $M_\phi = 256$.

Fifth, the Fourier transform can be calculated from $Z(\phi_j)$ by

$$\hat{Z}(\nu_k) = \frac{1}{N_\phi} \sum_{j=0}^{N_\phi} e^{-i\nu_k \phi_j} \cdot Z(\phi_j) \tag{4.8}$$

where $\nu_k = \dfrac{2\pi k}{\phi_{fin} - \phi_{ini}} = \dfrac{k}{M_T}$. Then, the pseudo-spectral function of concentration PA, $Z(\phi_j)$, is

$$S_Z(\nu_k) = |\hat{Z}(\nu_k)|^2. \tag{4.9}$$

Last, as we change the space from time t to pseudo-phase ϕ with a constant max-to-max spacing of 2π, the fundamental angular frequency of this ϕ-transformed concentration PA is $\nu_{M_T} = 1.0$. By summing up the pseudo-spectral at the fundamental angular frequency and all its "overtones", which are positive integer multiples of the fundamental angular frequency, we can get a quantity that is proportional to the square of the signal's amplitude. Consequently, the synchronization order parameter is

$$F_X = \sqrt{\frac{1}{S_Z(0)} \sum_{n=1}^{M_\phi/2} S_Z(\nu_{M_T \cdot n})}. \tag{4.10}$$

where $\nu_{M_T \cdot n}$, with $n = 1, 2, ..., M_\phi/2$, represents the fundamental angular frequency and all its "overtones", and $S_Z(0)$ is the pseudo-spectrum for the signal background.

Figure 4.9 provides a clear demonstration of the major process, from the second step to fifth step, for the generation of the synchronization order parameter, F_X. In summary, the basic idea of the synchronization order parameter is to extract the amplitude information from the PA signal by transforming the selected signal segment from the time space to a pseudo-phase space, where the max-to-max spacing is constant. As a result, the amplitude information of the PA signal can be easily accessed through its pseudo-spectral by using the built-in max-to-max frequency.

4.3 Phase transition

The foregoing results in Figures 3.8-3.11 show a rough idea that the qualitative behavior of the system can be changed from being oscillatory to non-oscillatory, by either keeping $C_4 = 0.9$ and decreasing Q, or keeping $Q = 1.0$ and decreasing C_4. With the help of the order parameter, F_X, we can quantitatively estimate the oscillatory behavior of the clock system as a function of the coupling parameters Q and C_4.

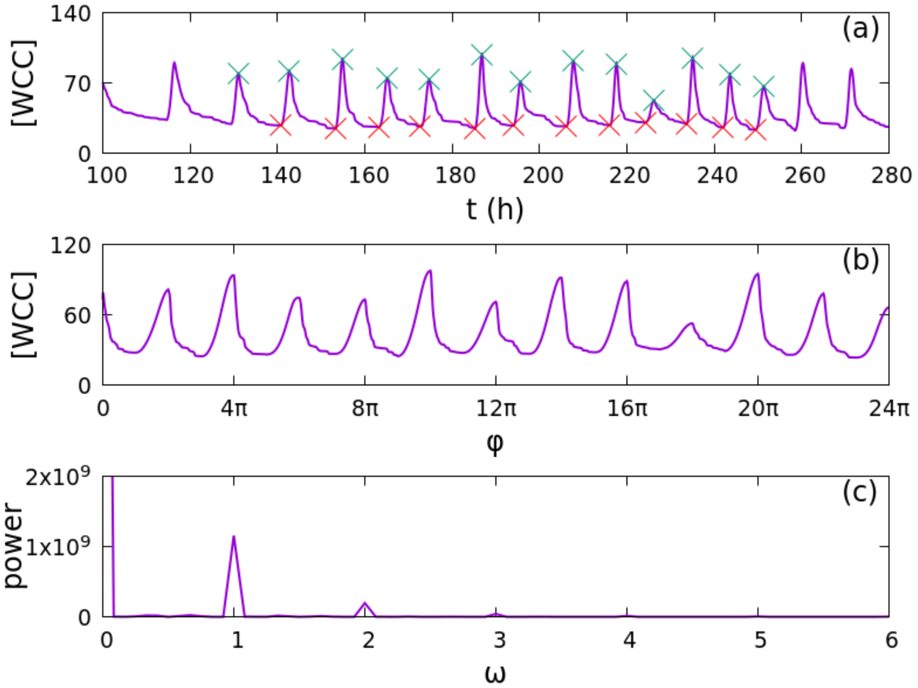

Figure 4.9: The generation process, from the second step to the fifth step, of synchronization order parameter, F_X. (a) A segment of $[WCC]_{PA,t}$ from a single random system trajectory, with green × representing DL maxima, t_m^+, and red × representing DL minima, t_m^-. (b) The selected $[WCC]_{PA,t}$ segment in the domain of pseudo-phase, ϕ (c) The pseudo-spectrum of the signal shown in (b).

Figures 4.10-4.11 show F_X of 4 species plotted vs. the coupling parameters Q and C_4, for finite population sizes N, along with $N \to \infty$ extrapolation, $F_X^{(\infty)}$. The mean and ± 1 SEM of each data point in the finite-N curves are estimated by bootstrapping a sample of $N_R = (8 \times 10^5)/N$ system trajectories.

To extrapolate [34] finite-N results to $F_X^{(\infty)}$, we first assume that the concentration PA, $[X]_{\mathrm{PA},N,t}$, for a population of N cells, is averaged over N signals, each of which contains noise, $Z_i \sim \mathcal{N}(0, \sigma)$. Thus $[X]_{\mathrm{PA},N,t}$ can be taken as a zero-mean variable with a standard deviation of $\sigma/\sqrt{N}$. Based on the definition of F_X, described in 4.2, both the signal relative amplitude information and noise information are able to be extracted from the pseudospectra of $[X]_{\text{PA},N,t}$. As a result, the relation between the signal relative amplitude and signal noise is shown below,

$$F_X(N) \sim F_X^{(\infty)} + g(\frac{1}{\sqrt{N}}), \tag{4.11}$$

where $g(\frac{1}{\sqrt{N}}) = 0$ as $N \to \infty$.

In Figures 4.10-4.11, the purple lines show the quadratic extrapolation [34] of $N = \infty$, where each point of $F_X^{(\infty)}$ is extrapolated from results of 5 different population sizes, by fitting them into Equation 4.11 with

$$g(\frac{1}{\sqrt{N}}) = A \cdot (\frac{1}{\sqrt{N}})^2 + B \cdot \frac{1}{\sqrt{N}}. \tag{4.12}$$

Figures 4.12-4.13 clearly show the well fitted lines for F_X of Q- and C_4-dependence, respectively, *vs.* $1/\sqrt{N}$.

From the definition of F_X, it is evident that when the coupling strength is not strong enough to regiment all cells' randomness, the PA signal of $N = \infty$ will become non-oscillatory and, consequently, $F_X^{(\infty)}$ will approach zero. This assertion can be observed in Figures 4.10-4.11, where $F_N^{(\infty)}$ is approximate equal to zero, for the four species, with either $Q \leq 0.86$ or $C_4 \leq 0.4$. These two figures also suggest a continuous phase transition from non-oscillatory ($F_X^{(\infty)} = 0$ to oscillatory dynamics ($F_X^{(\infty)}$), as a function of either Q or C_4.

To estimate the critical coupling strengths, Q_{crit} and $C_{4,crit}$, of these Q- and C_4-driven transitions, we fit the standard power law dependence to the $F_X^{(\infty)}$-data near the transition, in the form,

$$F_X^{(\infty)} = f \cdot |\theta|^{\beta} \cdot H(\theta), \text{ with } H(\theta) = \begin{cases} 1, \text{ for } \theta \geq 0 \\ 0, \text{ for } \theta < 0 \end{cases} \tag{4.13}$$

where $\theta := Q/Q_{crit} - 1$ or $\theta := C_4/C_{4,crit} - 1$. The comparisons of $F_X^{(\infty)}$-data and their non-linear fitted lines within ranges of [0.852,0.865] and [0.4, 0.5] for Q and $C4$, respectively, are shown in Figure 4.12 and Figure 4.13. The tabulated results in Tables 4.3-4.4 display agreements on the critical coupling strengths, Q_{critic} and $C_{4,critic}$, within statistical uncertainties, across four analyzed species, $X \equiv FRQ, WCC, CCG, Si$. Finally, it is very easy to observe that the critical coupling strengths, Q_{crit} and $C_{4,crit}$, are estimated within the range of [0.856,0.86] and [0.43,0.45], respectively, which agree with what are shown in Figures 4.10-4.11.

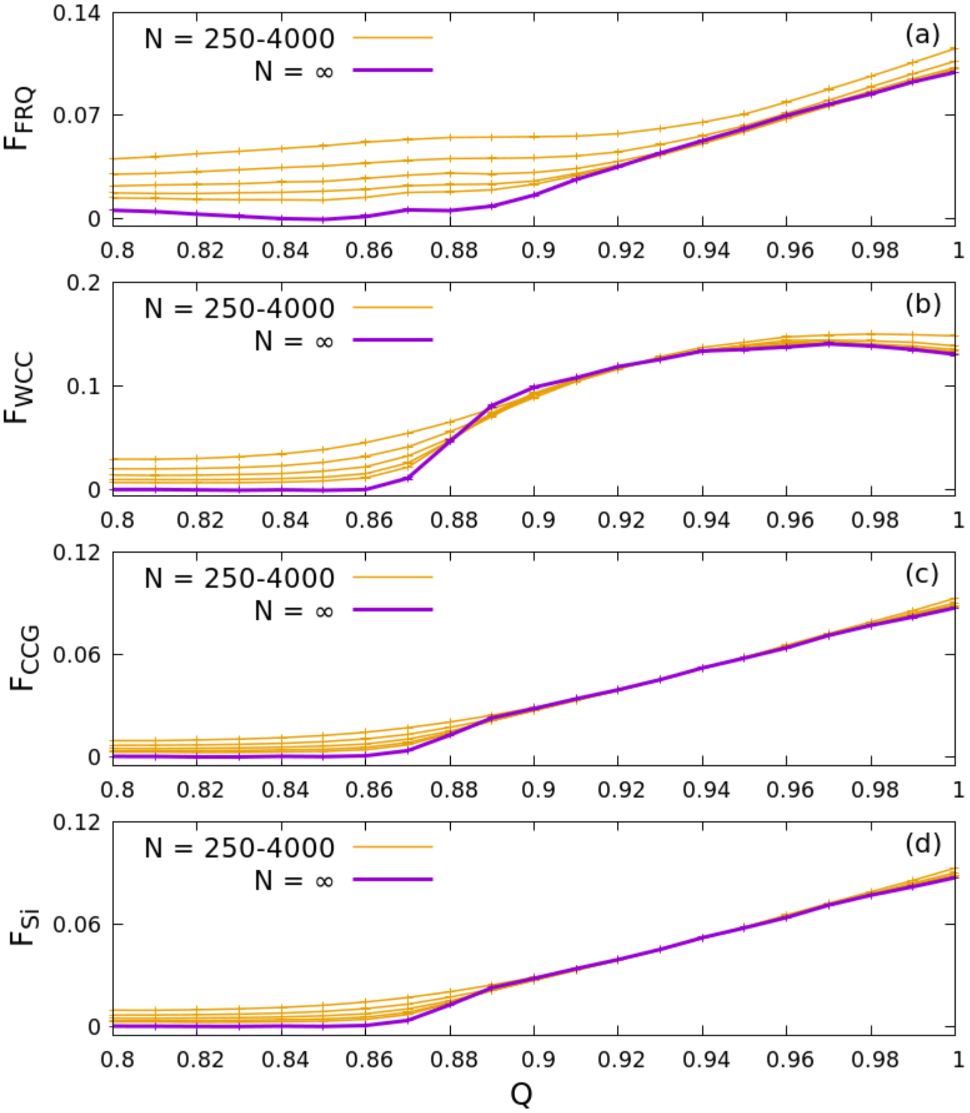

Figure 4.10: Synchronization order parameter, F_X, vs. extra-cell signal coupling parameter, Q, in population, with $C_4 = 0.9$, from $N = 250 - 4000$ and extrapolated to $N = \infty$. The population sizes of yellow lines, from top to bottom, are $N = 250, 500, 1000, 2000, 4000$ for $Q \leq 0.88$. The mean and ± 1 SEM of each data point in the finite-N curves are estimated by bootstrapping $B = 1000$ bootstrap samples from a base sample of $N_R = (8 \times 10^5)/N$ system trajectories for a N-cell population. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

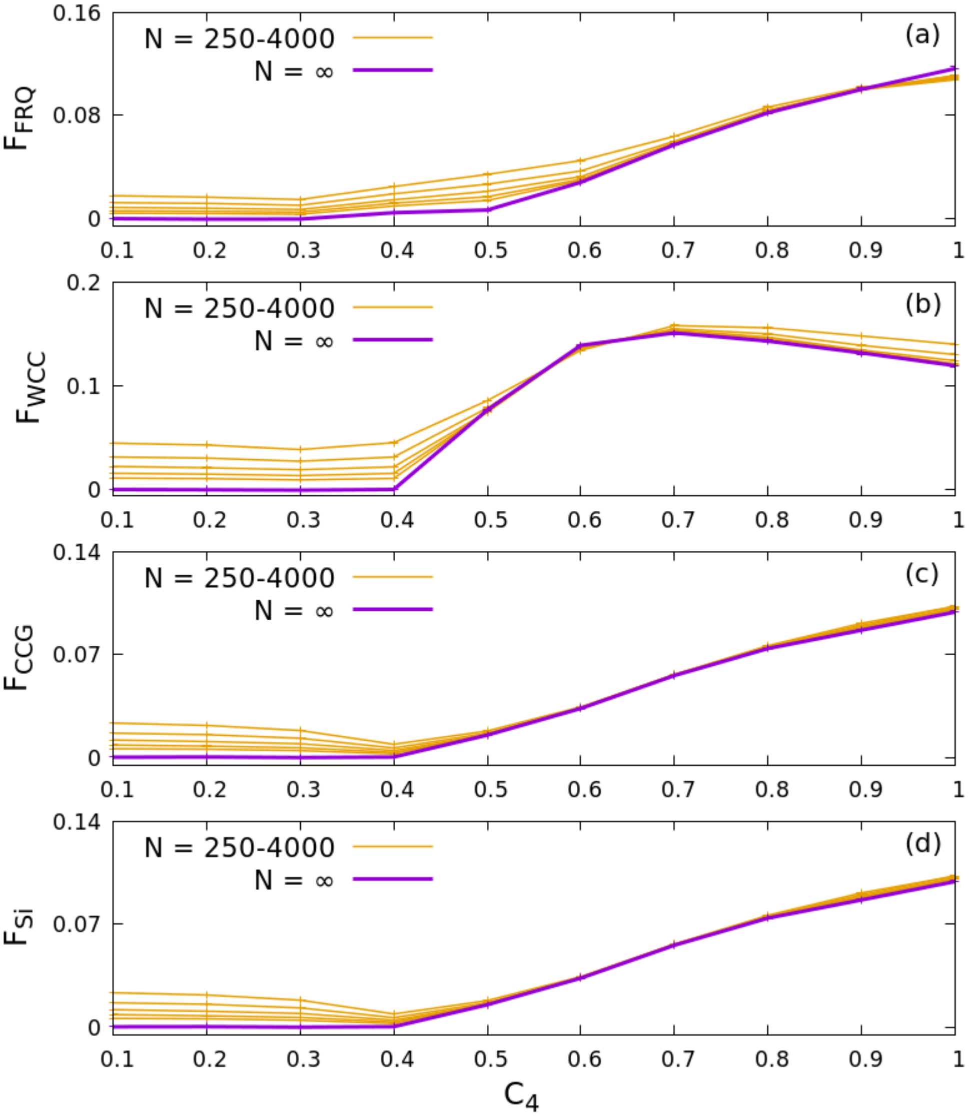

Figure 4.11: Synchronization order parameter, F_X, vs. extra-cell signal coupling parameter, C_4, in population, with $Q = 1.0$, from $N = 250 - 4000$ and extrapolated to $N = \infty$. The population sizes of yellow lines, from top to bottom, are $N = 250, 500, 1000, 2000, 4000$ for $C_4 \leq 0.5$. The mean and ± 1 SEM of each data point in the finite-N curves are estimated by bootstrapping $B = 1000$ bootstrap samples from a base sample of $N_R = (8 \times 10^5)/N$ system trajectories of a N-cell population. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

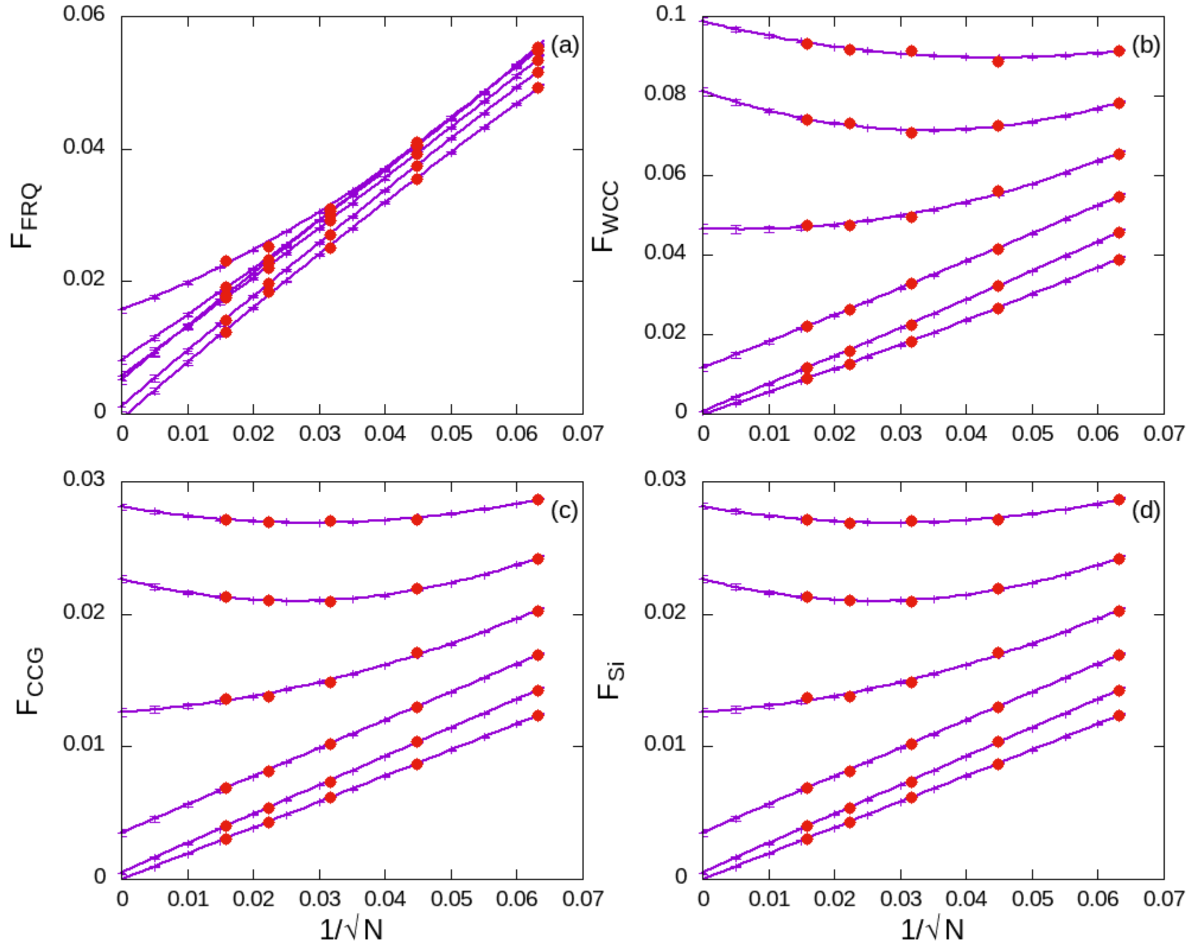

Figure 4.12: Synchronization order parameter, F_X, with $C_4 = 0.9$ and $Q = 0.85 - 0.9$, vs. $1/\sqrt{N}$. Red dots are simulation results of F_X with different Q and N. Purple lines are quadratic fitted lines for different values of Q, which, from top to bottom, are $Q = 0.85 - 0.9$. Means and ±1 SEM error bars for both F_X-data and fitted lines are estimated by bootstrapping $B = 1000$ bootstrap samples from base samples of $N_R = (8 \times 10^5)/N$ system trajectories for N-cell populations. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

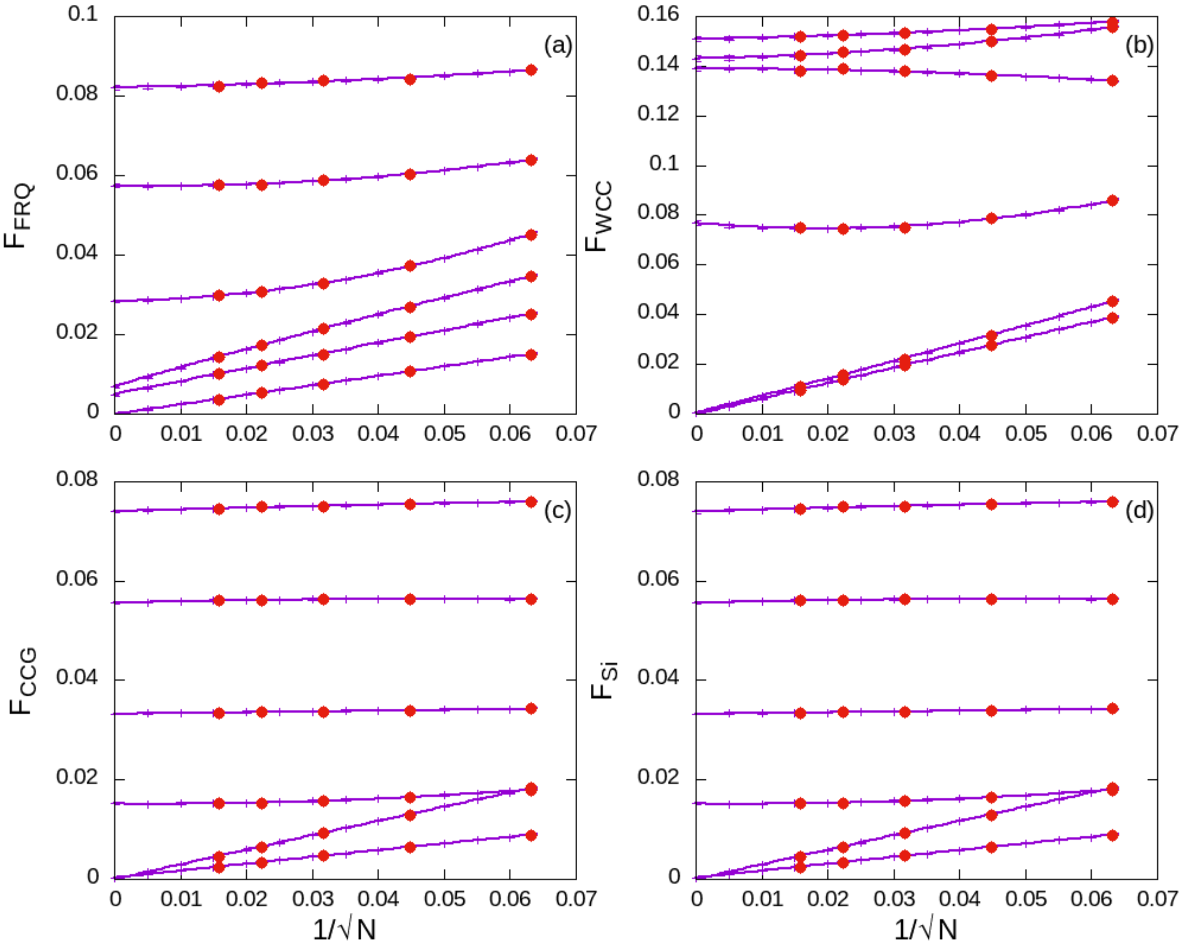

Figure 4.13: Synchronization order parameter, F_X, with $Q = 1.0$ and $C_4 = 0.3 - 0.8$, vs. $1/\sqrt{N}$. Red dots are simulation results of F_X with different C_4 and N. Purple lines are quadratic fitted lines for different values of C_4, which, from bottom to top, are $C_4 = 0.3 - 0.8$. Means and ±1 SEM error bars for both F_X-data and fitted lines are estimated by bootstrapping $B = 1000$ bootstrap samples from base samples of $N_R = (8 \times 10^5)/N$ system trajectories for N-cell populations. The error bars may be smaller than the plotted line width and therefore not visible in the plot.

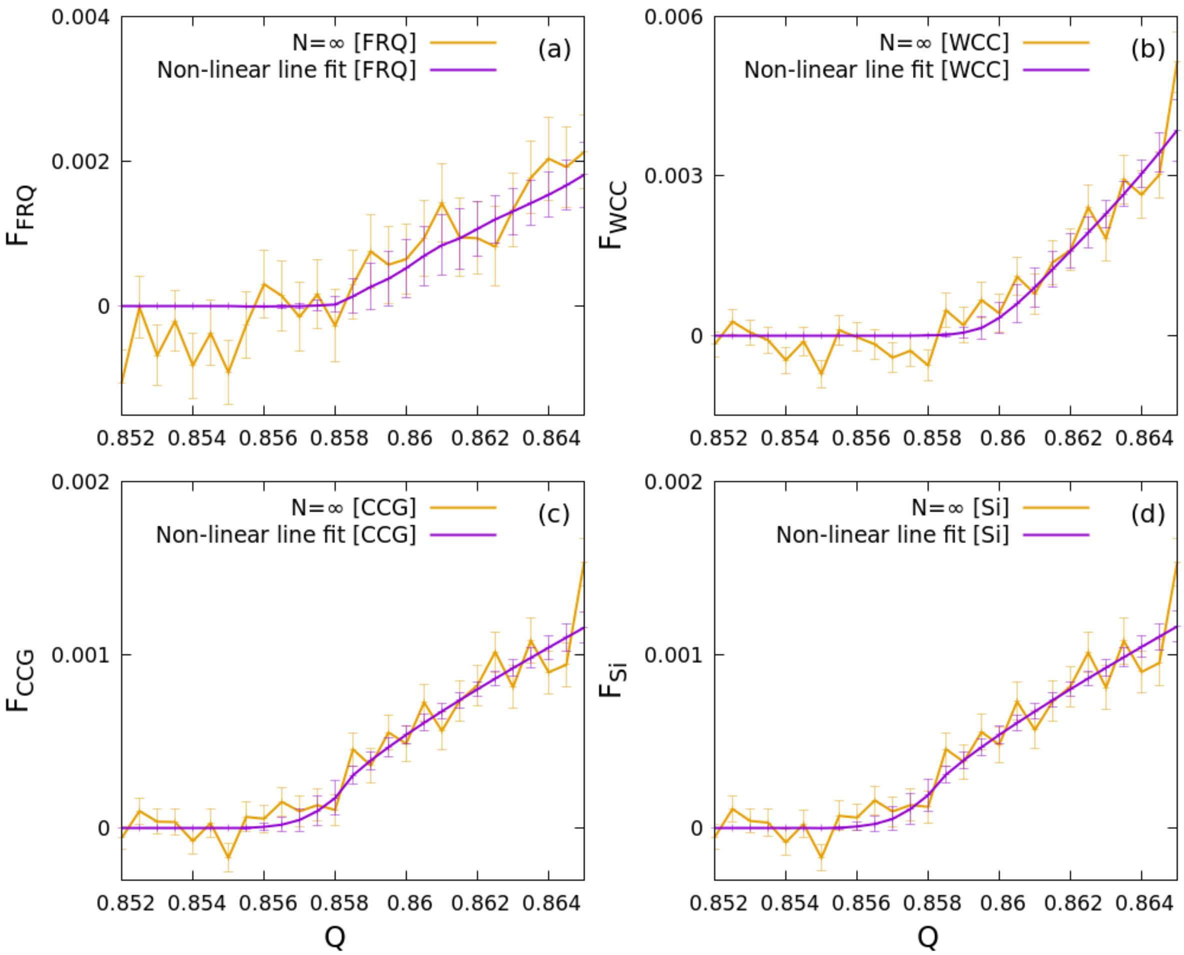

Figure 4.14: Synchronization order parameter, $F_X^{(\infty)}$, vs. Q in the vicinity of critical coupling strength, Q_{crit}. Yellow curves are the extrapolated $F_N^{(\infty)}$-data. Purple lines are standard power law dependence fitted lines for $F_X^{(\infty)}$-data. Means and ±1 SEM error bars for both $F_X^{(\infty)}$-data and fitted lines are estimated by bootstrapping $B = 1000$ bootstrap samples from base samples of $N_R = (1.6 \times 10^6)/N$ system trajectories for N-cell populations.

Table 4.3: Results for parameters of fitted curves to the $F_X^{(\infty)}$-data, with $X \equiv FRQ, WCC, CCG, Si$, around the Q-driven transition. Means and SEMs are obtained from bootstrapping $B = 1000$ bootstrap samples from a base sample of $N_R = (1.6 \times 10^6)/N$ system trajectories for N-cell populations.

	FRQ		WCC		CCG		Si	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Q_{crit}	0.8588	0.0014	0.8592	0.0010	0.8570	0.0009	0.8568	0.0009
β	0.9	1.3	1.26	0.78	0.80	0.24	0.82	0.25
f	0.2e+14	4.5e+14	0.04e+8	1.10e+8	0.08	0.11	0.2	1.3

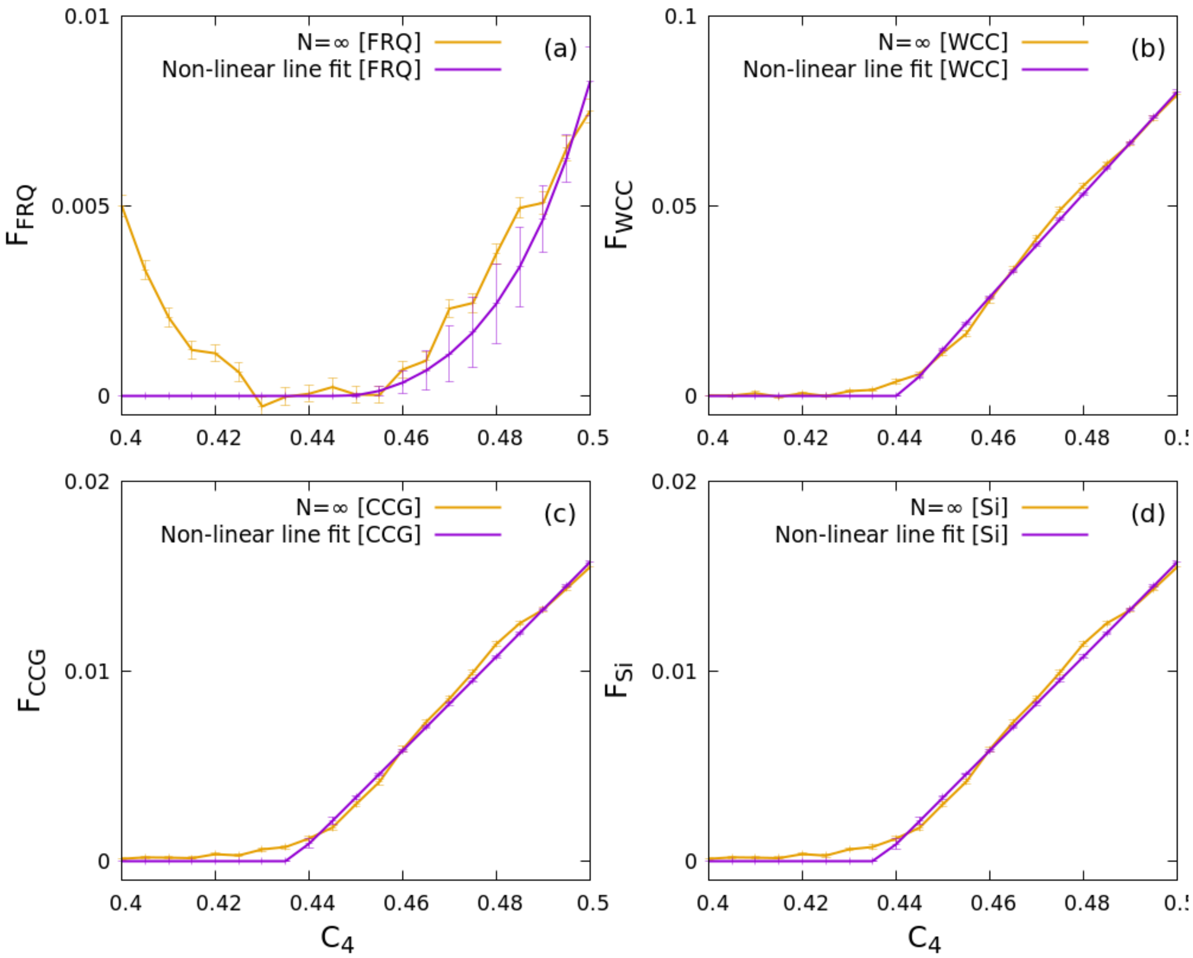

Figure 4.15: Synchronization order parameter, $F_X^{(\infty)}$, vs. C_4 in the vicinity of critical coupling strength, $C_{4,crit}$. Yellow curves are the extrapolated $F_N^{(\infty)}$-data. Purple lines are standard power law dependence fitted lines for $F_X^{(\infty)}$-data. Means and ±1 SEM error bars for both $F_X^{(\infty)}$-data and fitted lines are estimated by bootstrapping $B = 1000$ bootstrap samples from base samples of $N_R = (1.6 \times 10^6)/N$ system trajectories for N-cell population. Some error bars may be smaller than the plotted line width and therefore not visible in the plot.

Table 4.4: Results for parameters of fitted curves to the $F_X^{(\infty)}$-data, with $X \equiv FRQ, WCC, CCG, Si$, around the C_4-driven transition. Means and SEMs are obtained from bootstrapping $B = 1000$ bootstrap samples from base samples of $N_R = (1.6 \times 10^6)/N$ system trajectories for N-cell populations.

	FRQ		WCC		CCG		Si	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
$C_{4,crit}$	0.4463	0.0062	0.4415	0.0005	0.4361	0.0014	0.4362	0.0016
β	2.9	1.3	0.98	0.02	1.01	0.05	1.01	0.05
f	0.6e2	6e2	0.58	0.02	0.111	0.007	0.110	0.008

Chapter 5

Conclusion

In the work of this dissertation, we have presented a detailed investigation of a hybrid model, wherein stochastic gene regulations are combined with deterministic time evolution of species in a population of cells, for biological clock oscillators of the microbial fungus *N. crassa.* The quorum sensing-type coupling mechanism, used by the model for communication of cells, is based upon a mean field assumption that all signaling molecules can be instantaneously and uniformly distributed over the medium shared by the population. The simulation results of the multi-cell hybrid model present a seemingly paradoxical phenomenon that the cell population as a whole exhibits a coherent and approximately periodic oscillations, whereas individual cells in the population show completely irregular dynamics. The result remains almost the same for model with light exposures, except for increased number of gene flip events in each single-cell oscillator.

By means of the foregoing non-linear time series analysis methods, we can clearly describe the underlying physics for the paradoxical collective behavior of this model. We first use the *WinMax* filter approach to extract all the DL maxima from both PA and single-cell signals. Clock-readout results prove that single-cell clock oscillators, represented by four species, have either much slower running speeds or less stable period lengths, comparing with PA signals. From the results of beat skip probability and Pearson correlation, we can draw the conclusion that the quorum sensing signal randomly recruits only a small fraction of cells from the population as a sub-population to form each of the DL maxima in the PA signals. Furthermore, the membership of cells in these sub-populations are temporarily and changed rapidly on the time scales of typical oscillation periods of PA signals.

The robustness of system behaviors for the synchronization mechanism is explicitly exhibited by employing the synchronization order parameter, F_x, as a function of either the intra-cellular coupling parameter C_4 or the extra-cellular one, Q. Our results also suggest that a continuous phase transition exists in the infinite population limit, $N \to \infty$, as a function of C_4 or Q.

Bibliography

[1] M. Brunner and K. Káldi, Molecular microbiology **68**, 255 (2008).

[2] J. C. Dunlap, Cell **96**, 271 (1999).

[3] M. H. Vitaterna, J. S. Takahashi, and F. W. Turek, Alcohol Research & Health **25**, 85 (2001).

[4] C.-H. Chen and J. J. Loros, Communicative & Integrative Biology **2**, 448 (2009).

[5] W. Dong *et al.*, PloS one **3**, e3105 (2008).

[6] Z. Deng *et al.*, IEEE Access **7**, 49403 (2019).

[7] Y. Liu, N. Y. Garceau, J. J. Loros, and J. C. Dunlap, Cell **89**, 477 (1997).

[8] P. Ruoff, J. J. Loros, and J. C. Dunlap, Proceedings of the National Academy of Sciences **102**, 17681 (2005).

[9] H. Kitano, science **295**, 1662 (2002).

[10] A. Eldar and M. B. Elowitz, Nature **467**, 167 (2010).

[11] Y. Yu *et al.*, Proceedings of the National Academy of Sciences **104**, 2809 (2007).

[12] K. Lee, J. J. Loros, and J. C. Dunlap, Science **289**, 107 (2000).

[13] D. T. Gillespie, Journal of computational physics **22**, 403 (1976).

[14] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, Journal of Computational Physics **17**, 10 (1975).

[15] C. Caranica *et al.*, PloS one **13**, e0196435 (2018).

[16] M. Matsumoto and T. Nishimura, ACM Transactions on Modeling and Computer Simulation (TOMACS) **8**, 3 (1998).

[17] J. H. Cartwright and O. Piro, International Journal of Bifurcation and Chaos **2**, 427 (1992).

[18] R. Sedgewick and K. Wayne, *Algorithms*, 4st ed. (Addison-Wesley Professional, 2011).

[19] H. P. William, A. T. Saul, T. V. William, and P. F. Brian, *Numerical Recipes*, 3rd ed. (Cambridge University Press, 2007).

[20] R. W. Johnson, Teaching Statistics **23**, 49 (2001).

[21] E. Castro-Longoria, M. Ferry, S. Bartnicki-Garcia, J. Hasty, and S. Brody, Fungal Genetics and Biology **47**, 332 (2010).

[22] V. D. Gooch *et al.*, Eukaryotic cell **7**, 28 (2008).

[23] S. K. Crosthwaite, J. C. Dunlap, and J. J. Loros, Science **276**, 763 (1997).

[24] A. C. Froehlich, Y. Liu, J. J. Loros, and J. C. Dunlap, Science **297**, 815 (2002).

[25] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, Proceedings of the National Academy of Sciences **101**, 10955 (2004).

[26] M. B. Miller and B. L. Bassler, Annual Reviews in Microbiology **55**, 165 (2001).

[27] J. D. Dockery and J. P. Keener, Bulletin of mathematical biology **63**, 95 (2001).

[28] D. McMillen, N. Kopell, J. Hasty, and J. Collins, Proceedings of the National Academy of Sciences **99**, 679 (2002).

[29] H. P. William, A. T. Saul, T. V. William, and P. F. Brian, *Numerical Recipes*, 3rd ed. (Cambridge University Press, 2007).

[30] H. P. William, A. T. Saul, T. V. William, and P. F. Brian, *Numerical Recipes*, 3rd ed. (Cambridge University Press, 2007).

[31] T. D. V. Swinscow *et al.*, *Statistics at square one* (Bmj London, 2002).

[32] S. H. Strogatz, Physica D: Nonlinear Phenomena **143**, 1 (2000).

[33] T. Kreuz *et al.*, Physica D: Nonlinear Phenomena **225**, 29 (2007).

[34] H. P. William, A. T. Saul, T. V. William, and P. F. Brian, *Numerical Recipes*, 3rd ed. (Cambridge University Press, 2007).