How likely are oscillations in a genetic feedback loop with delay?

Filippo Cola

Department of Physics and Center for Complexity and Biosystems,

Università degli Studi di Milano and INFN,

via Celoria 16, 20133 Milano, Italy

Filippo Marchetti

Istituto di Chimica del Riconoscimento Molecolare,

Consiglio Nazionale delle Ricerche, via Mario Bianco 9, 20131 Milan, Italy

Guido Tiana[*]

Center for Complexity and Biosystems and Department of Physics,

Università degli Studi di Milano and INFN,

via Celoria 16, 20133 Milano, Italy

(Dated: August 7, 2017)

Abstract

Some genetic control networks display temporal oscillations as a result of delays in their homeostatic control. A relevant question about these systems is whether the oscillating regime is a rare feature, or it corresponds to a sizeable volume of the space of parameters. The answer is not trivial mainly due to the large number of parameters controlling the rate equations which describe the network. We have developed an efficient sampling scheme of the parameter space, based on a Monte Carlo algorithm, and applied it to a two-node system with delay, characterised by a 8-dimension parameter space. The result is that the volume fraction of parameter space associated with oscillations is small but not negligible, and it is weakly dependent on the duration of the delay. The most critical parameter to control oscillations is the coupling production rates, which must have opposite sign, giving rise to a negative feedback loop. The oscillating regions are connected except along the equilibrium constants between the two species, not allowing neutral evolution along this parameter.

---

[*]Electronic address: `guido.tiana@unimi.it`

I. INTRODUCTION

Several genetic control system in the cell have been observed to display regular temporal oscillations. This is the case, for example, of p53 [1, 2], Hes1 [3], NFκB [4] and Wnt [5]. As a rule, these oscillations have a period of the order of tens of minutes to hours, are associated to some specific mechanism and involve two main species linked in a feedback loop. P53 activates the transcription of Mdm2 which, in turn, promotes the degradation of p53; the resulting 20 minute oscillations are thought to coordinate the cellular response to DNA damage, causing growth arrest and eventually apoptosis [6]. Similarly, the oscillations of Hes1, which inhibits the transcription of its own mRNA, are turned into spatially repetitive units such as vertebrae, ribs and skeletal muscles during embryogenesis [7].

The simplest model to account for such oscillations is to describe the concentration of the species involved in the feedback loop making use of rate equations, characterised by a delay in the feedback. The delay summarises a set of molecular processes, which are not described explicitly in the equations, but which anyway take a finite time to be finalised. For example, in the case of the p53 loop, the expression of Mdm2 is activated by p53, but an increase in the transcription of p53 is not expected to lead instantaneously to the activation of Mdm2, due to the translocation, transcription and translation time associated with the process. The description of feedback loops with delay rate equations led to a fruitful characterisation of temporal oscillations for p53 [8], Hes1 [9], NFκB [10] and Wnt [11] (for reviews, see [12, 13]).

An important evolutionary question that one can put about these systems is how likely a feedback loop displaying some delay undergoes temporal oscillations. The dynamical behaviour of a feedback loop depends on a number of parameters, like dissociation constants, protein expression rates and so on. One would like to understand whether the emergence of oscillation requires that evolution carries out a careful fine–tuning of these parameters, or oscillations are a likely outcome of the dynamics of these systems.

To investigate this point, we studied a general two–species feedback loop controlled by 9 parameters. We set realistic ranges of variation for the parameters, which define a hyper–parallelepiped in the parameter space, and studied what fraction of the volume of the hyper–parallelepiped is occupied by oscillating dynamics, as compared to that occupied by dynamics which converges to a stationary state. In a two–dimensional system, chaotic behaviour is ruled out by Poincaré–Bendixson theorem.

2

Anyway, the systematic exploration of a 8–dimensional parameter space is rather cumbersome, and thus we developed a sampling scheme based on an optimisation principle. The idea is to define a function of the trajectory of the system which is minimised as oscillations appear, and to bias the exploration of parameter space with a Metropolis algorithm [14] using this function as an energy. After the sampling is finished, one estimates the volume of parameter space associated with oscillations by subtracting the effect of the bias, much similarly to what one does to obtain the density of states from a thermodynamic sampling.

The data can be analysed with some advanced tools developed in connection with molecular simulations, like the weighted–histogram method [15]. In this way one can quantify the fraction of the space of parameters associated with oscillations, and also estimate the effect of the different parameters in inducing them.

II.   THE FEEDBACK MODEL

A quite general description of a two-species biological feedback loop can be given by the delay rate equations

$$\frac{dx_1}{dt} = s_1 + \alpha_{12} \cdot \frac{x_2^h(t-\tau)}{k_{12}^h + x_2^h(t-\tau)} - \gamma_1 x_1$$
$$\frac{dx_2}{dt} = s_2 + \alpha_{21} \cdot \frac{x_1^h(t-\tau)}{k_{21}^h + x_1^h(t-\tau)} - \gamma_2 x_2, \qquad (1)$$

where x_1 and x_2 are the concentration of the two species, approximated as real numbers, s_1 and s_2 are their unconditioned production rates, α_{12} and α_{21} are their production rates dependent on the other species, γ_1 and γ_2 are their degradation rates. The coupling between the two species is provided by a delayed sigmoidal function, where the delay summarises a set of molecular processes which need some time to be completed, such as translocation, transcription and translation. For simplicity, it was assumed that these processes are similar for the two species and thus the same delay τ affects both of them. The sign of α_{12} and α_{21} determines if the relation between the two species is excitatory or inhibitory. We also set $h = 2$ since it is known from literature this is a common biological value for Hill's coefficient ([17]).

The 6 rates and the 2 equilibrium constants were allowed to vary in a range of values which we assume to be biologically relevant [18]. In particular, we set the constraints $10^{-10} \leq s_i \leq 100$ mol/s, $-100 \leq \alpha_{ij} \leq -10^{-10}$ and $10^{-10} \leq \alpha_{ij} < 100$ mol/s, $10^{-9} \leq k_{ij} < 10$ mol

and $10^{-8} \leq \gamma_i \leq 10 \text{ s}^{-1}$. We chose to express all times in seconds and all concentrations in mol.

The set of initial conditions $\{x_1(t), x_2(t)\}_{-\tau < t \leq 0}$ was set to the constant values $\{x_1^0, x_2^0\}$ obtained inserting a small perturbation ϵ in the fixed-point system of algebraic equations, that is

$$x_1^0 = \frac{\alpha_{12}}{\gamma_1} \cdot \frac{x_2^h}{x_2^h + k_{12}^h} + (1 + \epsilon)\frac{s_1}{\gamma_1}$$
$$x_2^0 = \frac{\alpha_{21}}{\gamma_2} \cdot \frac{x_1^h}{x_1^h + k_{21}^h} + (1 + \epsilon)\frac{s_2}{\gamma_2}. \tag{2}$$

If either x_1^0 or x_2^0 is negative, it is set to zero. In the present calculation we chose $\epsilon = 0.2$. Anyway, the fact that the functions at the right–hand–side of Eq. (1) are monotonic guarantees that the system displays a single fixed point. Consequently, the system converges to a stationary solution or oscillates independently on the initial point of the dynamics. In other words, the oscillating or stationary character is a property of the system controlled only by its parameters.

Equations (1) were solved with the adaptive–step Runge–Kutta method implemented in the gsl libraries [16]. The length of each simulation is 10^5s, corresponding in order of magnitude to the length of the cell cycle.

III. THERMODYNAMIC–LIKE SAMPLING OF THE SPACE OF PARAMETERS

Sampling the parameter space of the feedback model to identify the oscillating regions is not straightforward. An exhaustive analysis of all the possible set of parameters is unfeasible because of its high–dimensionality. Under the working hypothesis that the fraction of the space of parameters corresponding to an oscillating behaviour is small, a random sampling of the oscillating regions is not expected to be efficient. This problem is analogous to that of thermodynamic sampling, that is looking for the conformations of a complex system which have such a low energy to be statistically relevant according to Boltzmann distribution. In the thermodynamic problem, a purely random sampling is inefficient because the density of states is a non–decreasing function of energy, and consequently a random sampling would visit mainly high–energy states which are irrelevant according to Boltzmann statistics. To solve this problem, Monte Carlo methods have been introduced [14], which generate

4

a Markov chain that samples the space according to the very probability distribution one wants to recover, usually the Boltzmann distribution. In this way relevant states are visited more often than poorly–relevant ones.

Our strategy consists in sampling the parameter space of the model making use of a Markov chain much similar to that used by Monte Carlo methods. Given a set of parameters $\pi \equiv \{s_1, s_2, \alpha_{12}, \alpha_{21}, k_{12}, k_{21}, \gamma_1, \gamma_2\}$, we generated the solution trajectory as described in Sect. II, and associated to it a quantity Ω which counts how many steady oscillations the trajectory displayed. This is operatively defined as the number of maxima of $x_1(t)$ which assume, within a relative error of 1%, the same value of $x_1(t)$ (and consequently, after a transient, are not damped; cf. Fig. 1).

To favour the sampling of the high–Ω regions of the space of parameters, we started from a random set π chosen within the ranges defined in Sect. II and changed randomly the parameters, multiplying them by 10^r, where r is a random number such that $-1 < r < 1$. The multiplicative change of the parameters allows the system to explore efficiently all the orders of magnitude (i.e., effectively exploring the logarithm of the rates). In the case of α_{ij}, when the random move brings it above -10^{-12} or below 10^{-12}, the sign is changed; in this way the algorithm could sample continuously the two semi–regions in which the α_{ij} are allowed to move. The move was then accepted with a rate given by the Metropolis–like expression,

$$w(\pi \rightarrow \pi') = \min \left[1, \exp \left(\frac{\Omega(\pi') - \Omega(\pi)}{T} \right) \right]. \tag{3}$$

Here the parameter T is analogous to the temperature in a standard Metropolis simulation, and controls the bias towards high vaues of Ω. This procedure is essentially equivalent to a Metropolis simulation controlled by the energy $(-\Omega)$, where the minus causes the algorithm to maximise Ω instead of minimising it, as usually done with the energy.

Given Eq. (3), one can then use all the armoury of Monte Carlo algorithms to improve the sampling of the space of parameters and to analyse the data. For example, one can perform a parallel tempering sampling, in which different replica are sampled simultaneously at different temperatures, to help the system to jump through disconnected regions of non–zero values of Ω [20]. Actually, one can use all the advanced sampling techniques which were developed for enhancing the exploration of molecular conformations, such as simulated tempering, umbrella sampling, metadynamics, etc. [21].

Once the sampling reached convergence, the data were analysed with a weighted–

histogram method [15]. The main quantity of interest to quantify the fraction of space of parameters associated with oscillations is the density

$$g(\Omega) = \int d^8\pi \, \delta(\Omega(\pi) - \Omega), \tag{4}$$

which is analogous to the density of states of thermodynamic systems. From the histograms $N_T(\Omega)$ collected from the simulations at different temperatures, one can obtain the most likely form of $g(\Omega)$ compatible with them, obtaining [15]

$$g(\Omega) = \frac{\sum_T N_T(\Omega)}{\sum_T N_T \exp[\Omega/T]/Z_T} \tag{5}$$

where $N_T = \sum_\Omega N_T(\Omega)$ and the Z_T is obtained solving the set of implicit equations

$$\sum_\Omega \frac{\sum_{T'} N_{T'}(\Omega)}{\sum_{T'} N_{T'} \exp[\Omega/T' - \Omega/T] \cdot Z_T/Z_{T'}} = 1. \tag{6}$$

The function $g(\Omega)$ is proportional to the volume of the parameter space characterised by a given oscillation number Ω, independent on the bias used to make the sampling efficient.

IV. HOW LIKELY ARE OSCILLATIONS?

Simulations of the kind described in Sect. III were conducted for different choices of the delay ($\tau = 500, 1000, 1500, 2000, 2500$s), using each time three temperature parameters ($T = 1, 2, 3$). The simulations are computationally quite demanding, and took approximately 6 months of cpu time on Intel i7 3.4Ghz processors.

The resulting $g(\Omega)$ for different delays are displayed in Fig. 2 in semi–logarithmic scale. The points at $\Omega = 0$ and 1 indicate sets of parameters not associated with oscillations. At all values of the delay τ, $g(\Omega)$ displays an overall decreasing trend, compatibly with the idea that the stationary behaviour is the most common outcome, while oscillations are somewhat disfavoured. However, the strongest decrease occurs at $\Omega \lesssim 10$, corresponding to a oscillation frequency of less that 10^{-4} s^{-1}, while the decrease is only marginal at larger values of Ω. From the simulations it is difficult to assess if the wavy behaviour at large Ω is a real physical effect or noise associated with the finite statistics.

The volume fraction of the space of parameters associated with oscillating solutions of Eq. (1), defined by

$$f = \frac{\sum_{\Omega>1} g(\Omega)}{\sum_{\Omega\geq0} g(\Omega)}, \tag{7}$$

is displayed in Fig. 3 as a function of the delay τ. The volume fraction displays a spike at small delays ($\tau = 500$s), reaching a fraction of $f = 0.027$, and then converges to an almost constant value of $f = 0.004$. Perturbing the definition of f, for example defining as oscillating also the solutions displaying $\Omega = 1$ does not change the qualitative behaviour of this plot, suggesting that it is quite robust.

The effect of the 8 parameters of the model in inducing oscillations is displayed in Fig. 4, where the value of Ω sampled at the different values of the parameters of the rate equations are plotted for the Monte Carlo simulation generated at $\tau = 500$s. The simulations with the other delays are qualitatively similar (data not shown). Since the system of rate equations (1) is symmetric, we collected in the same plots the rates corresponding to the two species.

The dependence of oscillations on the mutual couplings α_{ij} ($i = 1, 2$) is bimodal, displaying large values of Ω at small (in absolute value) negative and at large positive values of α_{ij}. To further investigate this point, we plotted in Fig. 5 the dependence of Ω on the product $\alpha_{12} \cdot \alpha_{21}$. Oscillations ($\Omega > 1$) appears only if the product is negative and small. In other words, the equations must describe a negative feedback loop, in which one species activate the other, which, in turn, inhibits the former. Moreover, in the inset of Fig. 5 it is shown the maximum value of Ω as a function of the larger between α_{12} and α_{21}. The curve is approximately straight, indicating that to obtain oscillations with high frequency, the larger α_{ij} must be large and that such a relation is linear.

As a function of the equilibrium constants k_{ij}, the oscillation parameter Ω is rather variable, filling densely the scatter plot except for a region between 10^{-7} and 10^{-5} mol. Although one cannot rule out that the lack of points in this region is due to a bad sampling, the otherwise dense appearance of the plot, and the straight boundary between the populated and non–populated region may suggest the this region of parameter space is actually depleted of oscillating trajectories.

The scatter plot of Ω as a function of the degradation rates γ_i gathers all points with $\Omega > 1$ at $\gamma_i > 10^1 \text{s}^{-1}$. Note that the values of γ_i determine the time scale of the dynamics of $x_i(t)$ and, in absence of fixed ranges of variability of all the parameters, its value would be irrelevant, setting the time units of all the rates. The accumulation of the points with $\Omega > 1$ at the higher border of γ_i indicates that if it were allowed to be increased beyond this border, the system could find new oscillating regions by decreasing the other rates; this would anyway results in rates assuming unrealistic values when compared with the

underlying molecular processes.

The dependence of Ω on the constant sources s_i is straightforward, the system displaying oscillations only when both s_i are negligible. When this is not the case, the two equations are essentially decoupled, and the process becomes approximately a birth–death process, converging to a stationary solution.

V. DISCUSSION AND CONCLUSIONS

Oscillations in the concentrations of proteins are a response of the cell to external stimuli that has been observed in several cellular contexts [1–4, 22–24]. Even one of the simplest system displaying oscillations, that is a two–node delayed feedback loop, can be difficult to study because one has little control over the numerical values of the parameters which control the model. For most systems, only a qualitative accounting of the relation between genes is available [25], sometimes with the specification that the relation is excitatory or inhibitory, but the associated numbers are unknown [26]. When these are available, they are often extracted from biochemical or model experiments, and not directly related to the *in vivo* situation. Consequently, one is compelled to study the model by varying its parameters in a realistic, usually wide, range of values. However, this procedure is not straightforward even in the simple case of a two–node system, because this is defined by eight numerical parameters, and their exhaustive scan is unfeasible.

To face this problem, we have applied the sampling techniques and the analysis strategies originally developed to study molecular systems in the canonical ensemble. Analogously to the sampling of the space of parameters of out rate equations, molecular systems typically display a very high dimensionality, often reaching $10^5 - 10^6$ degrees of freedom. This suggests that these techniques can be effective also to study larger regulatory networks, controlled by many more parameters. In the case of molecular systems, sampling techniques build a Markov chain in conformational space to exploit energy correlations and thus find the low–energy conformations, which are the most rare but at the same time the most significant according to Boltzmann distribution. Similarly, we generated Markov chains in the space of the parameters of the delay equations, biasing the search for highly–oscillating solutions through a temperature–like parameter. Once we explored the available parameter space, using multiple temperatures to fasten the exploration, we removed the bias to obtain the

actual density of parameters associated with oscillations at various frequencies.

The result is that oscillations occupy a small, but not negligible volume (fraction of percent) of the entire parameter space. This fraction is quite independent on the delay, and increases only at small delay values (cf. Fig. 6). Oscillations appear only for negative feedback loops, within a rather well–defined range of the parameter which couples the two species, and only if their spontaneous production rate is negligible. All these data point to a scenario in which evolution had to search actively for oscillations. Although we already know that α_{12} and α_{21} must have opposite sign for the system to display oscillations [12], we let the system explore the whole parameter space to prove the generality of the approach and to constrain the sampling as little as possible; the constrain on the sign of this parameter eventually emerges as a result.

The projections of the oscillating regions of the space of parameters along the direction of the different rates are usually connected. Different seems to be the case along the direction of the equilibrium constants k_{ij}, which displays a gap spanning more than two orders of magnitude. This means that once evolution tunes the rates to produce oscillations, it is quite easy to adjust them, for example, to change their frequency. On the other hand, the model predicts an evolutionary barrier on the equilibrium constant which defines the interaction properties between the two species.

Author contribution statement: FC contributed to writing the calculation code, made the calculations and analyzed the results; FM wrote the calculation code; GT designed the strategy, contributed to analyze the results and wrote the article.

[1] Y. Haupt, R. Maya, A. Kazaz and M. Oren, Nature, **387**, 296 (1997).

[2] G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz and U. Alon, Nature Genetics **36**, 147 (2004).

[3] H. Hirata, S. Yoshiura, T. Ohtsuka, Y. S. Bessho, T. Harada, K. Yoshikawa and R. Kageyama, Science **298**, 840 (2002).

[4] A. Hoffman, A. Levchenko, M. L. Scott and D. Baltimore, Science **298**, 1241 (2002).

[5] M. L. Dequeant, E. Glynn, K. Gaudenz, M. Wahl and J. Chen, Science **314**, 1595 (2006).

[6] B. Vogelstein, D. Lane, A. J. Levine, Nature, **408** 307 (2000).

[7] O. Pourquie, J. Anat., **199**, 169 (2001).

[8] G. Tiana, M. H. Jensen and K. Sneppen, Europ. Phys. J. B **29**, 135 (2002).

[9] M. H. Jensen, K. Sneppen and G. Tiana, FEBS Letters **541**, 176 (2003).

[10] S. Krishna, M. H. Jensen and K. Sneppen, Proc. natl. Acad. Sci. USA **103**, 10840 (2006).

[11] P. B. Jensen PB, L. Pedersen, S. Krishna and M. H. Jensen, Biophys. J. **98**, 943 92010).

[12] G. Tiana, S. Krishna S. Pigolotti, M. H. Jensen and K. Sneppen, Physical Biology **4**, R1 (2007).

[13] G. Tiana and M. H. Jensen, Phil. Trans. R. Soc. A **371**, 20120469 (2013).

[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[15] A. Ferrenberg and R. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[16] M. Galassi et al., *GNU Scientific Library Reference Manual*, 3rd Ed. (Network Theory Ltd, London 2009).

[17] J. P. J. Hetherington, A. Warner and R. M. Seymour, J. R. Soc. Interface, **3**, 319 (2006).

[18] B. Schwanhusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach, Nature, **473**, 337 (2011).

[19] J. Ferkinghoff-Borg, Eur. Phys. J. B **29**, 481 (2002).

[20] R. Swendsen and J. Wang, Phys. Rev. Lett. **57**, 2607 (1986).

[21] D. Frenkel, *Understanding Molecular Simulations*, 2nd Edition (Academic Press, San Diego, 2002)

[22] C. Simon and G. brandenberger, Diabetes **51**, S258 (2002)

[23] F. Vasile, E. Pechkova and C. Nicolini, Proteins **70**, 1112 (2008)

[24] T. Ikegami, Y. Maruyama, H. Doi, A. Hattori and H. Ando, Front. Neurosci. **9**, 9 (2015)

[25] H. Han *et al.*, Sci. Rep. **5**, 11432 (2015).

[26] M. Ronen, R. Rosemberg, B. I. Shraiman and U. Alon, Proc. Natl. Acad. Sci. USA **99**, 10555 (2002).
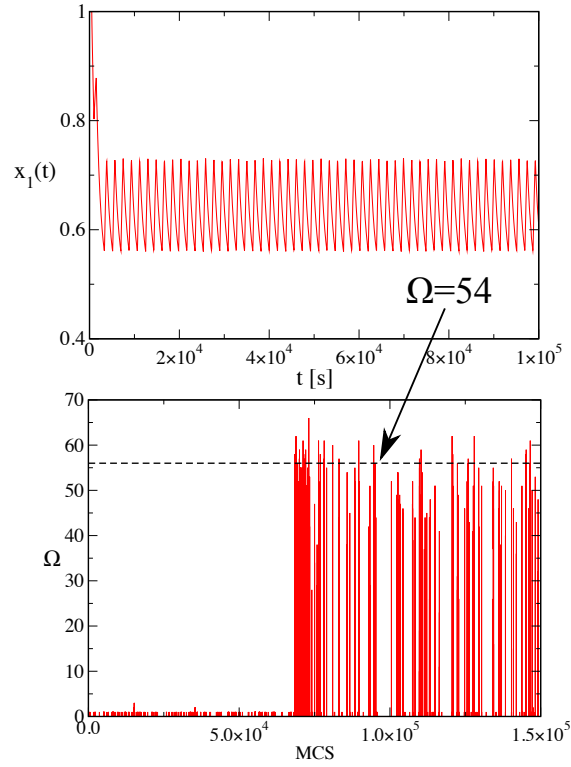
FIG. 1: (Upper panel) An example of oscillating solution $x_1(t)$ of the feedback loop, displaying $\Omega = 54$. (Lower panel) An example of Monte Carlo sampling of the space of parameters, where the system first visits a region corresponding to stationary solutions ($\Omega \leq 1$) and later finds oscillating solutions.
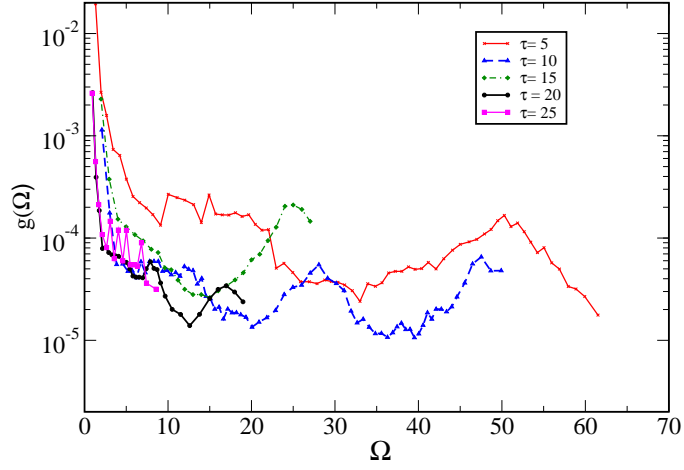
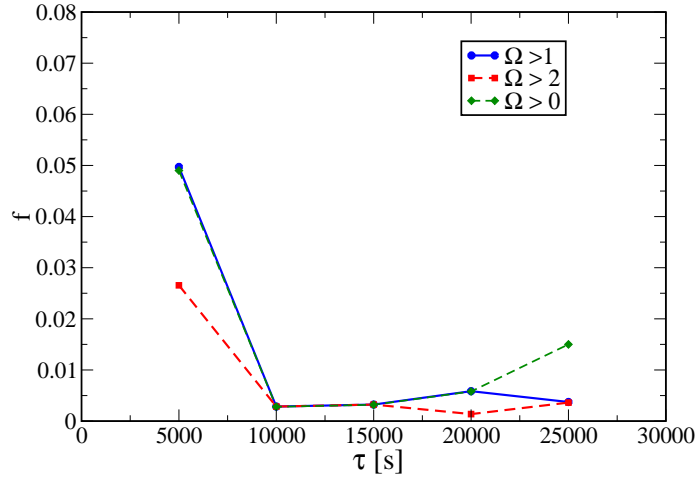FIG. 2: The density $g(\Omega)$ calculated for dfferent values of the delay τ.

FIG. 3: The fraction f of space of parameters associated with fluctuating trajectories, defined as $\Omega > 1$, as a function of the delay τ (circles). To test the robustness of the result, we perturbed the definition of fluctuating trajectories, calculating the behaviour of f at $\Omega > 2$ (squares) and $\Omega > 0$ (diamonds).
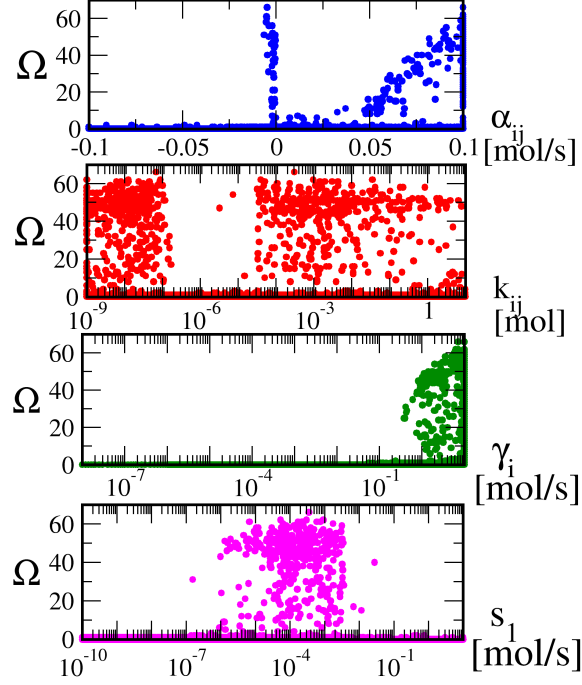
12

FIG. 4: The values of the number of oscillations Ω associated with the points of parameter space sampled by the simulation carried out at $\tau = 500s$ and $T = 1$ projected, respectively, onto the parameters α_{ij}, k_{ij}, γ_i and s_i. To make the results more readable, α_{ij} is plotted in a linear scale, while the other parameters in logarithmic scale. The density of points in each region of the plots reflects the probability that the delay equations display a specific value of Ω for the value of the parameters that define that region.
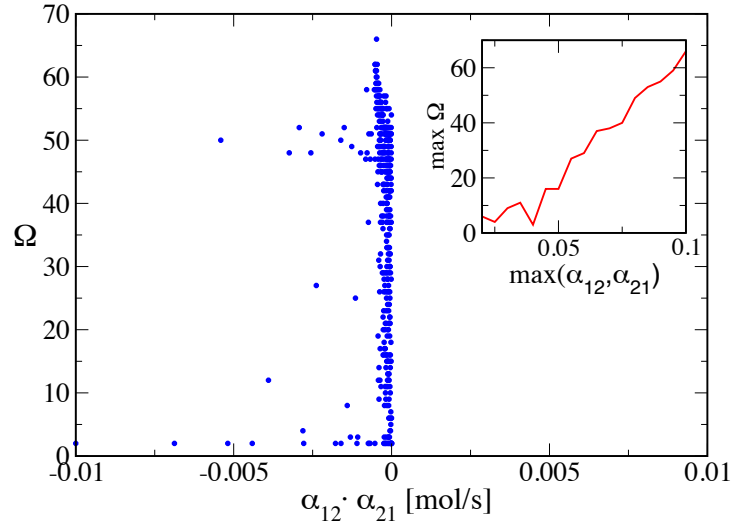
FIG. 5: The values of Ω sampled by the simulation carried out at $\tau = 500s$ and $T = 1$ as a function of the product between the α_{ij} of the two species. In the inset, the maximum value of Ω as a function of the larger between α_{12} and α_{21}.

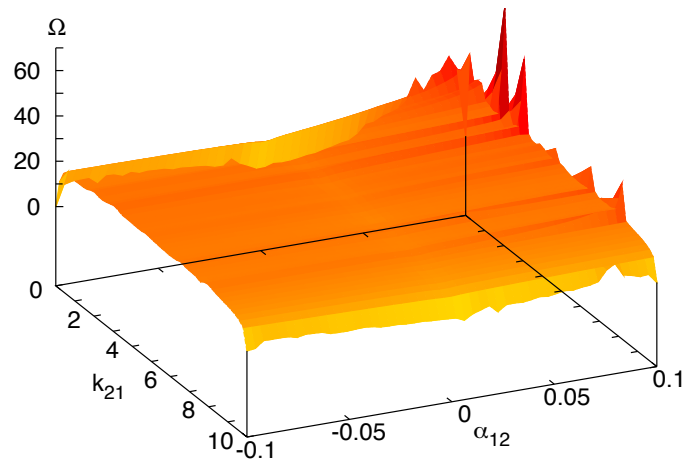FIG. 6: The values of Ω sampled as a function of α_{12} and k_{21} suggest that the oscillating region is a small but not negligible part of the parameter space.

14