Stochastic genetic networks with solvable structures

Conference Paper · December 2014

DOI: 10.1063/1.4904627

CITATIONS
2

READS
41

1 author:

Ovidiu Lipan
University of Richmond
54 PUBLICATIONS **3,046** CITATIONS

SEE PROFILE

Stochastic Genetic Networks with Solvable Structures

Ovidiu Lipan

Department of Physics, University of Richmond, Richmond, Virginia 23173, U.S.A.

Abstract. We describe a set of basic stochastic biocircuits for which the Master Equation is completely solvable. Beside linear circuits, which are known to be solvable, we show that tree-like circuits with polynomial transition functions are also completely solvable. We associate a simple but unambiguous graphical representation to such circuits. The graphical representation shows the signal propagation through these simple circuits.

Keywords: stochastic genetic networks
PACS: 87.18.Cf,87.15.H-,87.15.Ya

INTRODUCTION

This study is driven by three main goals: (i) the need to express the regulatory genetic networks as visual diagrams (ii) the mathematical model for a genetic network should be stochastic and unambiguously recoverable from the visual diagram, and (iii) the stochastic mathematical model should be completely solvable. The first goal stems from the usual presentation of the data in the field of molecular biology [1, 2]. Pathway maps of molecular interactions drive the scientific dialog and also represent the source of new testable hypothesis for the experimentalist. The second goal asks for a stochastic model instead of a deterministic one [3, 4, 5, 6]. Although deterministic models are useful for some genetic networks, the basic fact that a molecular interaction is a probabilistic phenomenon implies that the fundamental mathematical model should incorporate stochastic principles. The second goal also require a tight correspondence between the visual diagram and the mathematical model. Such a correspondence, setting aside the sytochasticity, presents an interest by itself [7].

The study of solvable models, which is the third goal, is valuable at least for two reasons. One reason is that for solvable models there is no need to use numerical simulations which can be costly for large networks. A second reason is that solvable models can be used to check different analytic approximation methods for unsolvable models and also to develop heuristics for efficient algorithms.

THE MATHEMATICAL MODEL AND ITS SOLUTION FOR A TREE-LIKE STOCHASTIC BIOCIRCUIT

In what follow we present, by way of example, a short introduction to the notations we use for the visual diagrams and the associated stochastic model. An extended description can be found in [3]. These notations will also be used to describe the solvable models. The example consists of a biocircuit composed of three molecules Fig.1A. Two molecules interact to regulate the production of the third molecule. In chemical reaction notation the process is codified by $A+B\to C$. The molecule A binds to the molecule B forming a complex $A:B$ which then drives the production of the third molecule C. The complex $A:B$ is not present as an additional molecule in Fig.1A. The probability per unit time for the production of the molecule C is proportional with the product q_1q_2, [8]. Here q_1 and q_2 represent the number of molecules of type A and B, respectively. The number of molecules of the C-type is denoted by q_3. The set of numbers $q=(q_1,q_2,q_3)$ represent the state of the biocircuit. The set is composed of nonegative integers, so the model is discrete in the state variable. The state changes in time and fluctuates stochastically. Both characteristics are specified through the probability, $P(q_1,q_2,q_3,t)$, that the number of molecules takes the values (q_1,q_2,q_3) at time t. This probability, written in short form as $P(q,t)$, is known only after solving the Master Equation:

$$\frac{\partial P(q,t)}{\partial t}=\sum_{\varepsilon}T_{\varepsilon}(q-\varepsilon,t)P(q-\varepsilon,t)-P(q,t)\sum_{\varepsilon}T_{\varepsilon}(q,t) \tag{1}$$

10th International Conference on Mathematical Problems in Engineering, Aerospace and Sciences
AIP Conf. Proc. 1637, 582-589 (2014); doi: 10.1063/1.4904627

Here, the initial probability distribution $P(q,0)$ is given, together with the transition probabilities per unit time $T_\varepsilon(q,t)$. The transition probabilities represent the building blocks of the stochastic model and are based on experimental data or they are considered as part of the definition of the biocircuit. For the biocircuit from Fig.1A, the ε variables are $\varepsilon_1 = (1,0,0), \varepsilon_2 = (0,1,0)$ and $\varepsilon_3 = (0,0,1)$, with the corresponding transition probabilities $T_{\varepsilon_1}(q,t) = g_1(t)$, $T_{\varepsilon_2}(q,t) = g_2(t)$ and $T_{\varepsilon_3}(q,t) = f(t)q_1q_2$. The difference $q-\varepsilon_3$ in (1) is (q_1,q_2,q_3-1) and similarly for the other two transitions. To see the meaning of this difference consider that the biocircuit is in the state (q_1,q_2,q_3-1) at some time t. Then the biocircuit jumps to (q_1,q_2,q_3), through an increase of q_3 by one molecule. The increase is due to T_{ε_3}. The lines that start at q_1 and q_2 join together to represent the product q_1q_2 and than ends on the box representing ε_3. These lines describe the formula for the transition probability T_{ε_3} that controls the behavior of the third molecule. For this reason the lines that represent the transition probabilities are called control lines. From the other types of lines, that start from the center of the boxes and end on molecules, we can read the components of the vector ε. For ε_3 there is only one line that ends on q_3 so only the third component of ε_3 is not zero. If another line were to start from the box representing ε_3 and end on q_2 than that ε_3 would be $(0,1,1)$. These lines are called action lines. The arrow at the end of an action line depicts, as is customary in biology, a production of a molecule, thus a positive number for the corresponding component of ε. A negative number is depicted by ending the line with a short bar. The control lines that act on ε_1 and ε_2 can be imagined as starting from other molecules that have a constant number equal to 1. Such constant molecule numbers are not represented in the molecular diagram. In what follows we generalize the transition probabilities that appeared in the reaction $A+B \to C$. The generalization is based on the system of decreasing factorials

$$e_{m_k} = q_k(q_k-1)...(q_k-m_k+1) \tag{2}$$

where k labels each molecule present in the biocircuit. In a system with q_k molecules of specie k, the probability for a collision of m_k molecules (all of the same species k) is $M^{m_k}(t)e_{m_k}$, [8]. The superscript m_k for M^{m_k} must be generalized to a vector $\overline{m} = (m_1,m_2,...,m_N)$ if interactions between N different molecular species are considered. The bar over m is needed to distinguish the vector index $\overline{m}$ from a tensor index $m = (11122333...)$ that contains the number 1 m_1 times, the number 2 m_2 times and so on. The tensor index will be used later.

The transition probability, for one action ε, is represented as a linear combination of products of decreasing factorials

$$T_\varepsilon(q,t) = \sum_{\overline{m}} M_\varepsilon^{\overline{m}}(t)e_{\overline{m}}(q) \tag{3}$$

with $e_{\overline{m}}(q) = e_{m_1}(q_1)e_{m_2}(q_2)....$ In a biocircuit, for a given ε, the majority of the functions $M_\varepsilon^{\overline{m}}$ are zero, except for those represented by the control lines that end on the box corresponding to the given ε. To the set of decreasing factorials we add $e_0 = 1$ which represent the control line coming from a molecule that does not change in time, like the control lines $g_1(t)$ and $g_2(t)$ in Fig.1A. The set of all decreasing factorial forms a basis of polynomials that is used to represent the transition probabilities for the time evolution of the biocircuit. The experimentally measured value usually consists of the mean and standard deviations of the molecule numbers $q_k, k = 1...N$. These statistical quantities can be obtained from the partial derivatives at $z \equiv (z_1,...,z_N)$ of the generating function

$$F(z_1,z_2,...,z_N,t) = \sum_{q_1=0,...,q_N=0}^{\infty} z_1^{q_1}...z_N^{q_N}P(q_1,...,q_N,t) \tag{4}$$

For example, the mean value of q_3 is $\partial_{z_3}F$ computed at $(z_1 = 1,...,z_N = 1)$. The standard deviation of q_3 can be computed from the mean value of $q_3(q_3-1)$ which is the second derivative $\partial_{z_3,z_3}F$ evaluated at $z = 1$. The correlations are obtained from mixed partial derivatives, like $\partial_{z_1,z_2}F$. All these partial derivatives evaluated at $z = 1$ are called factorial moments. The tensor index come in handy in representing these factorial moments: $F_m(t) = \partial_{\overline{m}}F(z,t)$ computed at $z = 1$. The vector $\overline{m}$ and the tensor index m will be used interchangeably. The Master Equation (1) transforms into an equation for the generating function

$$\frac{\partial F}{\partial t} = \sum_{\varepsilon,m}(z^\varepsilon - 1)z^{\overline{m}}M_\varepsilon^m\partial_m F \tag{5}$$

Here the vector z raised at a vector power ε or $\overline{m}$ means the product of each z-component raised at the corresponding component of the vector power, $z_1^{m_1}z_2^{m_2}...z_N^{m_N}$. A system of ordinary differential equations for the time evolution of the factorial moments $F_m(t)$ is obtained from (5). For this we introduce the function $Q_m(\varepsilon)$ defined by $\partial_m z^\varepsilon = Q_m(\varepsilon)z^{\varepsilon-\overline{m}}$, where the partial derivative is taken componentwise $\partial_m = \partial_{m_1}\partial_{m_2}....$ The tensor index is useful for indexing the

moments $F_m(t)$ because taking a partial derivative ∂_α over z_α extends the index m to $m\alpha$. The Greek letters $\alpha, \beta, ...$ are used to denote a tensor index of length 1. The length of an index is $m_1 + ... + m_N$. If desirable, after concatenation, the tensor index $m\alpha$ can be rearranged in an increasing order.

The time evolution of the factorial moments are obtained from applying the operator $\partial_{\alpha_1,\alpha_2...}$ on the Master Equation (5). The action of this operator on a product of function $f(z)g(z)$ is written as

$$\partial_{\alpha_1...\alpha_n}(fg) = \{\partial_{\alpha_1...\alpha_k} f \partial_{\alpha_{k+1}...\alpha_n} g\}_\alpha \tag{6}$$

The braces indicates a summation over all pairs of disjoint sets $\alpha_1, ..., \alpha_k$ and $\alpha_{k+1}, ..., \alpha_n$. Finally, the time evolution of the factorial moments is

$$\frac{d}{dt} F_{\alpha_1...\alpha_n}(t) = \sum_m \{R^m_{\alpha_1...\alpha_k}(t) F_{m\alpha_{k+1}...\alpha_n}(t)\}_\alpha \tag{7}$$

The functions $R^m_{\alpha_1...\alpha_k}(t)$ are linear combinations over the actions ε of the controls $M^m_\varepsilon(t)$

$$R^m_{\alpha_1...\alpha_k}(t) = \sum_\varepsilon C^\varepsilon_{\alpha_1...\alpha_k}(m) M^m_\varepsilon(t) \tag{8}$$

The time-independent coefficients $C^\varepsilon_{\alpha_1...\alpha_k}(m)$ are sums of products of the functions $Q_{\alpha_k}(\overline{m}), Q_{\alpha_k}(\varepsilon + \overline{m})$, $k = 1...n$. For example

$$R^m_{\alpha_1\alpha_2}(t) = \sum_\varepsilon (Q_{\alpha_1\alpha_2}(\varepsilon + \overline{m}) - Q_{\alpha_1\alpha_2}(\overline{m})) M^m_\varepsilon(t) \tag{9}$$

where the Q-function with concatenated indices $\alpha_1\alpha_2$ is recurrently obtained from the definition $Q_{m\alpha}(\varepsilon) = Q_m(\varepsilon) Q_\alpha(\varepsilon - \overline{m})$.

Usually, for a biocircuit composed of N molecular species only the mean values F_{α_i} and the second order moments $F_{\alpha_i\alpha_j}$, $i, j = 1...N$ are of interest. Unfortunately, for a general biocircuit, the system of equations (7) does not produce a finite closed subsystem of equations that involve only the first and the second order moments. For example, the existence of feedback loops in the biocircuit produce moments of order 3 and possible higher on the right hand side of the (7) computed for $n = 2$. The concatenated tensor index $m\alpha_1...\alpha_n$ in (7) is responsible for increasing the order of the moments on the right side of (7) with respect to its left side. For linear biocircuits, for which the transition probabilities T_ε are linear functions of q_k, $k = 1..N$, the functions $R^m_{\alpha_1...\alpha_k}(t)$ are conveniently zero in such a way that the equations for the first, second and all other moment order close exactly at that moment order.

So, the question is if there are nonlinear biocircuits for which a closed finite subsystem of equations can be extracted from (7). For example, it may be the case that to obtain the time evolution of $F_{\alpha_i\alpha_j}$ we need also to find the time evolution of a 4-th order moment $F_{\beta_i\beta_j\beta_k\beta_l}$. However, the system of equation involving some additional higher order moments should be finite.

The answer is positive. In fact all tree-like biocircuits with polynomial transition probabilities (3) are solvable, producing a finite system of equations for any desired moment order.

By a tree-like biocircuit we understand a network that does not contain any loops. The structure is a directed graph connecting "parents" with "children". For a tree-like biocircuit, a term $(z^\varepsilon - 1) z^{\overline{m}} \partial_m F$ describes the control of parents m on the children ε. In other words, if a molecule q_k is present in the parent index $\overline{m} = (m_1, m_2, ..., m_n)$, that is $m_k \neq 0$, then the molecule q_k is not part of the children $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, which means $\varepsilon_k = 0$. The reverse is also true, namely if $\varepsilon_j \neq 0$ then $m_j = 0$.

Consider now a molecule q_p that belongs to the parent set that act on the children set from which we will select one molecule q_c.

We want to compute dF_c/dt and dF_{cc}/dt and show that a finite system of equations can be formed for the time evolution of F_c and F_{cc}. Using $\partial_c z^{\overline{m}} = 0$ and $\partial_p z^\varepsilon = 0$ we get the contribution (symbolized by $\leftarrow$) of $(z^\varepsilon - 1) z^{\overline{m}} \partial_m F$ to $\partial_c \partial_t F$

$$\begin{aligned}
\partial_c \partial_t F \leftarrow & Q_c(\varepsilon) z^{\varepsilon - \overline{c}} z^{\overline{m}} \partial_m F + (z^\varepsilon - 1) z^{\overline{m}} \partial_{mc} F \\
\partial_{cc} \partial_t F \leftarrow & Q_c(\varepsilon) Q_c(\varepsilon - \overline{c}) z^{\varepsilon - 2\overline{c}} z^{\overline{m}} \partial_m F + \\
& + 2 Q_c(\varepsilon) z^{\varepsilon - \overline{c}} z^{\overline{m}} \partial_{mc} F + (z^\varepsilon - 1) z^{\overline{m}} \partial_{mcc} F
\end{aligned}$$

Evaluating the result at $z = 1$ the term containing F_{mcc} is eliminated but F_{mc} remains in the system. Because of the concatenation with the tensor index m, the order mc is higher than 2, which may lead to an infinite system of

equations. Fortunately the time evolution of F_{mc} depends only on the parent index m, a fact that leads to a finite system of equations. Indeed, take a parent molecule q_p from the control index m and find

$$\partial_p \partial_c \partial_t F \leftarrow Q_c(\varepsilon) z^{\varepsilon - \overline{c}} \partial_p (z^{\overline{m}} \partial_m F) + (z^{\varepsilon} - 1) \partial_p (z^{\overline{m}} \partial_{mc} F) \tag{10}$$

The F_{mc} disappears form the left hand side of (10) when it is evaluated at $z = 1$. Thus the time evolution of F_c, F_{cc}, F_{mc} can be written only in terms of the moments of parent indices. The system of equation will contain moments higher than 2, but all these higher moments refer to parent indices. The same logic applies for F_{ccc} and higher children moments.

To solve a tree-like structure we need to decide first on the order of the highest moment for the molecules that does not control any other ones. These molecules are found at the bottom of the tree. Going backwards from the bottom of the tree towards its top, the system of equations will move from one children layer to a parent layer until the equations for the molecules on the top of the tree are found. The needed equations are selected step by step as the computation proceeds upward in the tree. The last equations are provided by the molecules from the top of the tree. Looking at the structure of the finite system of equations, the molecules from the top of the tree will contribute with moments of order higher than those below them. Going down in the tree, the moment order contributed by the molecules decreases, until the bottom of the tree is reached. There the moment order is the one which was set from the beginning.

To exemplify the procedure, we will study the simple but fundamental biocircuit of Fig.1A. The finite system of equations that goes up to second order moment for the third molecule, F_{33}, contains 13 equations. The moment F_{33} depends only on F_{123}. The parent molecules q_1 and q_2 contribute with a moment of order 4, F_{1122}. This is needed to find the correlation of all three molecules F_{123}.

1. $F_1' = g_1$
2. $F_{11}' = 2g_1F_1$
3. $F_2' = g_2$
4. $F_{22}' = 2g_2F_2$
5. $F_{12}' = g_2F_1 + g_1F_2$
6. $F_3' = fF_{12}$
7. $F_{112}' = g_2F_{11} + 2g_1F_{12}$
8. $F_{122}' = 2g_2F_{12} + g_1F_{22}$
9. $F_{1122}' = 2g_2F_{112} + 2g_1F_{122}$
10. $F_{13}' = fF_{112} + fF_{12} + g_1F_3$
11. $F_{23}' = fF_{12} + fF_{122} + g_2F_3$
12. $F_{123}' = fF_{112} + fF_{1122} + fF_{12} + fF_{122} + g_2F_{13} + g_1F_{23}$
13. $F_{33}' = 2fF_{123}$

The stochastic state of the biocircuit is (q_1, q_2, q_3). Focusing on the system of moments, the state can be viewed as a set composed of 13 moments that form a finite system of equations. The increase from 3 to 13 in the number of state-components is the trade-off for avoiding generating tens of thousands of different stochastic trajectories for the stochastic state (q_1, q_2, q_3) in order to compute the evolution of the moments. The system of equations for the moments opens the possibility of treating the biocircuit evolution as a linear input-output mapping. For example, a simple input-output relation become apparent for the mean values of q_1 and q_2

$$\begin{aligned} F_1(t) &= F_1^0 + \int_0^t dt_1 g_1(t_1) \\ F_2(t) &= F_2^0 + \int_0^t dt_1 g_2(t_1) \end{aligned} \tag{11}$$

Here F_1^0, F_2^0 are the initial mean values $F_1(0), F_2(0)$ and are considered as the input variables. The output variables are $F_1(t), F_2(t)$.

The linear input-output relation for all the other moments can be represented as nested integrals. For example, the time evolution of the mean value for the q_3 molecule is

$$\begin{aligned} F_3(t) = &\int_0^t dt_3 f(t_3) \int_0^{t_3} dt_2\, g_1(t_2) \int_0^{t_2} dt_1\, g_2(t_1) + \int_0^t dt_3 f(t_3) \int_0^{t_3} dt_2\, g_2(t_2) \int_0^{t_2} dt_1\, g_1(t_1) + \\ &F_1^0 \int_0^t dt_3 f(t_3) \int_0^{t_3} dt_2\, g_2(t_2) + F_2^0 \int_0^t dt_3 f(t_3) \int_0^{t_3} dt_2\, g_1(t_2) + \\ &F_{12}^0 \int_0^t dt_3 f(t_3) + F_3^0 \end{aligned} \tag{12}$$

As Fig.1A shows, the product of the molecules $q_1 q_2$ controls of F_3. To make this product visible in $F_3(t)$ we use

$$\begin{aligned}&\int_0^{t_3} dt_2\, g_1(t_2) \int_0^{t_2} dt_1\, g_2(t_1) + \int_0^{t_3} dt_2\, g_2(t_2) \int_0^{t_2} dt_1\, g_1(t_1) = \\ &\left(\int_0^{t_3} dt_2\, g_1(t_2)\right)\left(\int_0^{t_3} dt_2\, g_2(t_2)\right)\end{aligned} \tag{13}$$

and obtain

$$\begin{aligned}F_3(t) = &\int_0^t dt_3 f(t_3) \left(\int_0^{t_3} dt_2\, g_1(t_2)\right)\left(\int_0^{t_3} dt_2\, g_2(t_2)\right) + \\ &F_1^0 \int_0^t dt_3 f(t_3) \int_0^{t_3} dt_2\, g_2(t_2) + F_2^0 \int_0^t dt_3 f(t_3) \int_0^{t_3} dt_2\, g_1(t_2) + \\ &F_{12}^0 \int_0^{t_4} dt_3 f(t_3) + F_3^0\end{aligned} \tag{14}$$

Dropping the integral sign in a nested integral we arrive at a simple notation for the mean value $F_3(t)$

$$F_3(t) = f g_1 g_2 + f g_2 g_1 + f g_2 F_1^0 + f g_1 F_2^0 + f F_{12}^0 + F_3^0 \tag{15}$$

Representing the product rule (13) as $g_1 g_2 + g_2 g_1 = (g_1 g_2)$ we get

$$F_3(t) = f(g_1 g_2) + f g_2 F_1^0 + f g_1 F_2^0 + f F_{12}^0 + F_3^0 \tag{16}$$

The moments with a zero superscript represent, as before, the values the moments take at the initial time and are considered the input variables.The output variable is $F_3(t)$. The time evolution can be represented in a diagram form. Each integral is represented by a wiggly line labeled by the function that is integrated. A nested integral is represented by connecting a sequence of wiggly lines. The connector is a circle with a dark rim. The term $f g_1 g_2$ is represented by a three wiggly lines Fig.1B. For the product of two integrals the connector is a dark disk. The diagram for the free term $f g_1 g_2 + f g_2 g_1 = f(g_1 g_2)$ in (15) and (16) is presented in Fig.1B. The product form of the nested integral closely resembles the molecular diagram. The nested integrals diagrams can easily be extended to incorporate the initial conditions F^0, using, for example strait instead of wiggly lines. For this paper the nested integrals diagrams are restricted to the free terms which are terms that does not depend on any initial moments.

$F_{12}(t)$, which represents the average of the product $q_1 q_2$, is especially important for this biocircuit because the third molecule q_3 is controlled by the product of the molecule numbers q_1 and q_2. The solution for $F_{12}(t)$ is

$$F_{12}(t) = (g_1 g_2) + g_1 F_2^0 + g_2 F_1^0 + F_{12}^0 \tag{17}$$

The mean value $F_3(t)$, (16), is obtained by gluing the transition probability function $f(t)$ from the left of (17). The gluing process applies for other moments also, Fig.2A, where the free term of F_{112} is obtained by gluing the free term of F_1 on every node of the diagram representing the free term of F_{12}.

The free term of $F_{33}(t)$ is

$$\begin{aligned}&4fffg_1g_2g_1g_2 + 4fffg_2g_1g_1g_2 + 4fffg_1g_2g_2g_1 + 4fffg_2g_1g_2g_1 + 4fffg_1g_1g_2g_2 + 4fffg_2g_2g_1g_1 + \\ &2ffg_2fg_1g_2g_1 + 2ffg_2fg_1g_1g_2 + 2ffg_2fg_2g_1g_1 + 2ffg_1fg_1g_2g_2 + 2ffg_1fg_2g_1g_2 + 2ffg_1fg_2g_2g_1 + \\ &fg_1g_2fg_1g_2 + fg_2g_1fg_1g_2 + fg_1g_2fg_2g_1 + fg_2g_1fg_2g_1 + \\ &2fffg_1g_2g_1 + 2fffg_2g_1g_1 + 2fffg_1g_2g_2 + 2fffg_2g_1g_2 + 2fffg_1g_1g_2 + 2fffg_2g_2g_1 + \\ &fg_1fg_1g_2 + fg_1fg_2g_1 + fg_2fg_1g_2 + fg_2fg_2g_1 + fffg_1g_2 + fffg_2g_1\end{aligned} \tag{18}$$

Each nested integral factor of (18) can be represented by diagrams. Some of these terms are drawn in Fig.2A. These diagrams can also be interpreted as being generated by gluing together two smaller diagrams.

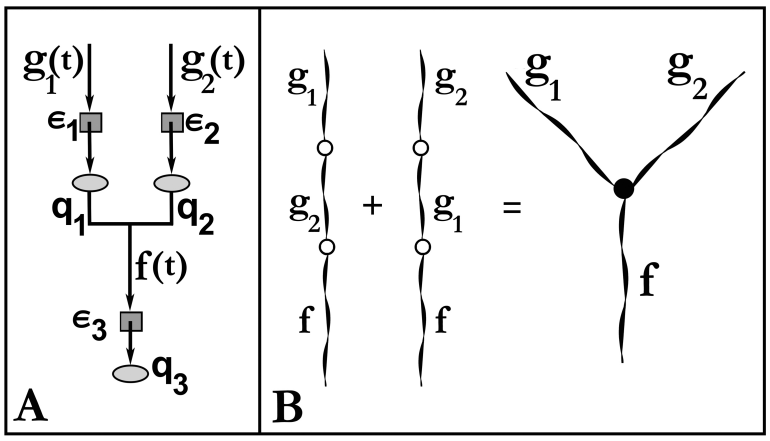

FIGURE 1. The basic control $A+B \to C$. A) The A,B and C molecule numbers are q_1, q_2 and q_3 respectively. B) The nested integrals diagrams for the free term of the mean value F_3. The diagram in the product form follows closely the molecular diagram from panel A.

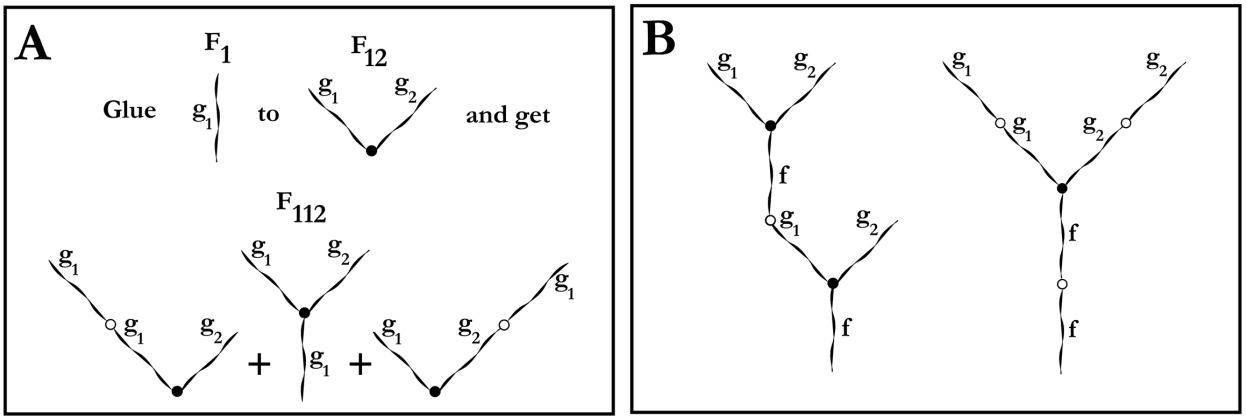

FIGURE 2. Gluing the nested integrals diagrams. A) The free term of the moment F_{112} can be obtained from gluing the diagrams of the free terms that correspond to F_1 and F_{12}. It can also be obtained by gluing F_{11} with F_2, not shown. B) Larger diagrams, with 6 branches, are necessary to represent some terms from F_{33}. Other terms need fewer branches.

CONCLUSION

The factorial moments for tree-like biocircuits with polynomial transition functions form a closed system of ordinary differential equations at any desired moment order. The formulas above connect the initial-time moments F^0, seen as inputs, with later times values $F(t)$, seen as outputs. The input-output relation is linear having time-dependent coefficients. The coefficients are nested integrals of functions that multiply the decreasing factorials in the transition probabilities.

Appendix: The input-output relation for the moments of the biocircuit from Fig.1A

Below we present the output values $F(t)$ as a function of the input variables $F(t=0) \equiv F^0$. The coefficient that does not depend on any F^0 is called *free*. If the coefficient is zero, the corresponding input variable F^0 is not present in the list. The $F_{33}(t)$ moment is obtained by gluing the transition function $f(t)$ to the left of F_{123}.

1. **$\mathbf{F_1}$**
   $free$: g_1
   F_1^0: 1
2. **$\mathbf{F_2}$**
   $free$: g_2
   F_2^0: 1
3. **$\mathbf{F_{11}}$**
   $free$: $2g_1g_1$
   F_1^0: $2g_1$
   F_{11}^0: 1
4. **$\mathbf{F_{22}}$**
   $free$: $2g_2g_2$
   F_2^0: $2g_2$
   F_{22}^0: 1
5. **$\mathbf{F_{12}}$**
   $free$: $g_1g_2+g_2g_1$
   F_1^0: g_2
   F_2^0: g_1
   F_{12}^0: 1
6. **$\mathbf{F_3}$**
   $free$: $fg_1g_2+fg_2g_1$
   F_1^0: fg_2
   F_2^0: fg_1
   F_{12}^0: f
   F_3^0: 1
7. **$\mathbf{F_{112}}$**
   $free$: $2g_1g_1g_2 + 2g_1g_2g_1 + 2g_2g_1g_1$
   F_1^0: $2g_1g_2+2g_2g_1$
   F_2^0: $2g_1g_1$
   F_{11}^0: g_2
   F_{12}^0: $2g_1$
   F_{112}^0: 1
8. **$\mathbf{F_{122}}$**
   $free$: $2g_2g_2g_1 + 2g_2g_1g_2 + 2g_1g_2g_2$
   F_1^0: $2g_2g_2$
   F_2^0: $2g_1g_2+2g_2g_1$
   F_{22}^0: g_1
   F_{12}^0: $2g_2$
   F_{122}^0: 1
9. **$\mathbf{F_{1122}}$**
   $free$: $4g_1g_2g_1g_2 + 4g_2g_1g_1g_2 + 4g_1g_2g_2g_1 + 4g_2g_1g_2g_1 + 4g_1g_1g_2g_2+4g_2g_2g_1g_1$
   F_1^0: $4g_1g_2g_2+4g_2g_1g_2+4g_2g_2g_1$
   F_2^0: $4g_2g_1g_1+4g_1g_2g_1+4g_1g_1g_2$
   F_{11}^0: $2g_2g_2$
   F_{22}^0: $2g_1g_1$
   F_{12}^0: $4g_1g_2+4g_2g_1$
   F_{112}^0: $2g_2$
   F_{122}^0: $2g_1$
   F_{1122}^0: 1

$\mathbf{F_{13}}$

$free$: $2fg_1g_2g_1+2fg_1g_1g_2+2fg_2g_1g_1+g_1fg_1g_2+g_1fg_2g_1+fg_1g_2+fg_2g_1$
F_1^0: $2fg_2g_1+2fg_1g_2+g_1fg_2+fg_2$
F_2^0: $g_1fg_1+2fg_1g_1+fg_1$
F_3^0: g_1
F_{11}^0: fg_2
F_{12}^0: $2fg_1+g_1f+f$
F_{13}^0: 1
F_{112}^0: f

$\mathbf{F_{23}}$

$free$: $2fg_1g_2g_2+2fg_2g_1g_2+2fg_2g_2g_1+g_2fg_1g_2+g_2fg_2g_1+fg_1g_2+fg_2g_1$
F_1^0: $2fg_2g_2+g_2fg_2+fg_2$
F_2^0: $2fg_1g_2+2fg_2g_1+g_2fg_1+fg_1$
F_3^0: g_2
F_{22}^0: fg_1
F_{12}^0: $2fg_2+g_2f+f$
F_{23}^0: 1
F_{122}^0: f

$\mathbf{F_{123}}$

$free$: $4fg_1g_2g_1g_2 + 4fg_2g_1g_1g_2 + 4fg_1g_2g_2g_1 + 4fg_2g_1g_2g_1 + 4fg_1g_1g_2g_2 + 4fg_2g_2g_1g_1 + 2g_2fg_1g_2g_1 + 2g_2fg_1g_1g_2+2g_2fg_2g_1g_1+2g_1fg_1g_2g_2+2g_1fg_2g_1g_2+2g_1fg_2g_2g_1+g_1g_2fg_1g_2+g_2g_1fg_1g_2+g_1g_2fg_2g_1+g_2g_1fg_2g_1+2fg_1g_2g_1+2fg_2g_1g_1+2fg_1g_2g_2+2fg_2g_1g_2+2fg_1g_1g_2+2fg_2g_2g_1+g_1fg_1g_2+g_1fg_2g_1+g_2fg_1g_2+g_2fg_2g_1+fg_1g_2+fg_2g_1$

F_1^0: $4fg_1g_2g_2+4fg_2g_1g_2+4fg_2g_2g_1+2g_1fg_2g_2+2g_2fg_1g_2+2g_2fg_2g_1+g_1g_2fg_2+g_2g_1fg_2+2fg_2g_1+2fg_1g_2+2fg_2g_2+g_1fg_2+g_2fg_2+fg_2$

F_2^0: $4fg_2g_1g_1+4fg_1g_2g_1+4fg_1g_1g_2+2g_2fg_1g_1+2g_1fg_2g_1+2g_1fg_1g_2+g_2g_1fg_1+g_1g_2fg_1+2fg_1g_2+2fg_2g_1+2fg_1g_1+g_2fg_1+g_1fg_1+fg_1$

F_3^0: $g_1g_2+g_2g_1$

F_{11}^0: $2fg_2g_2+g_2fg_2+fg_2$

F_{22}^0: $2fg_1g_1+g_1fg_1+fg_1$

F_{12}^0: $4fg_1g_2+4fg_2g_1+2g_1fg_2+2g_2fg_1+g_1g_2f+g_2g_1f+2fg_1+2fg_2+g_2f+g_1f+f$

F_{13}^0: g_2

F_{23}^0: g_1

F_{112}^0: $2fg_2+g_2f+f$

F_{122}^0: $2fg_1+g_1f+f$

F_{1122}^0: f

F_{123}^0: 1

REFERENCES

1. D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio **42**, D472–D477 (2014).
2. D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, and C. v. Mering **39**, D561–D568 (2011).
3. S. Achimescu, and O. Lipan, *IEE Systems Biology* **153 Issue:3**, 120 – 134 (2006).
4. O. Lipan, J. M. Navenot, Z. Wang, L. Huang, and S. C. Peiper, *PLoS Computational Biology* **3(10)**, 1859–11870 (2007).
5. F. P. Kelly, *Reversibility and Stochastic Networks*, 2011.
6. A. S. Ribeiro, and J. Lloyd-Price **23**, 777–779 (2007).
7. F. Büchel, N. Rodriguez, N. Swainston, C. Wrzodek, T. Czauderna, R. Keller, F. Mittag, M. Schubert, M. Glont, M. Golebiewski, M. van Iersel, S. Keating, M. Rall, M. Wybrow, H. Hermjakob, M. Hucka, D. B Kell, W. Müller, P. Mendes, A. Zell, C. Chaouiya, J. Saez-Rodriguez, F. Schreiber, C. Laibe, A. Dräger, and N. Le Novère, *BMC Systems Biology* **7**, 116 (2013).
8. van Kampen, N.G., *Stochastic Processes in Physics and Chemistry*, North Holland, Amsterdam, 1992.