Universidade de São Paulo

Instituto de Física de São Carlos

Guilherme da Costa Pereira Innocentini

Uma generalização do modelo de spins e bósons para a transcrição de genes sob múltiplo controle

São Carlos

2012

Guilherme da Costa Pereira Innocentini

Uma generalização do modelo de spins e bósons para a transcrição de genes sob múltiplo controle

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de concentração: Física Básica.

Orientador: Prof. Dr. José Eduardo Martinho Hornos.

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Innocentini, Guilherme da Costa Pereira
Uma generalização do modelo de spins e bósons para a transcrição de genes sob múltiplo controle / Guilherme da Costa Pereira Innocentini; orientador José Eduardo Martinho Hornos - versão corrigida -- São Carlos, 2012.
77 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Básica) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2012.

1. Expressão gência. 2. Regulação transcricional. 3. Estocasticidade. 4. Soluções exatas. I. Martinho Hornos, José Eduardo, orient. II. Título.

AGRADECIMENTOS

Agradeço aos professores e funcionários do Instituto de Física de São Carlos onde realizei este trabalho. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) que apoiou financeiramente este trabalho. Aos meus pais, Lourenço Innocentini Neto e Cristina Helena da C.P. Innocentini, que sempre estiveram amorosamente ao meu lado, me apoiando e me educando. À Emilie Tarbagayre pelo companheirismo, o carinho, a paciência e, também, pela rica e deliciosa introdução do farci, tajine, bavette, confit de canard, etc. Aos meus amigos Marcel Orlandi Paiano, Helber Canelhas, Diogo Oliveira Soares-Pinto, William Couto Corrêa de Lima, pelas boas e inspiradoras conversas que tivemos, em geral acompanhadas de um bom café ou um drink, ligeiramente mais forte e sempre com moderação, na saudosa e querida cidade maravilhosa, vulgo Rio de Janeiro. Ao Prof. Daniel Turolla Vanzella sempre claro, objetivo e criativo em suas exposições. Por último, ao meu orientador Prof. José Eduardo Martinho Hornos, pela orientação, pelos ensinamentos e, acima de tudo, pela amizade e confiança.

RESUMO

INNOCENTINI G.C.P. **Uma generalização do modelo de spins e bósons para a transcrição de genes sob múltiplo controle**. 2012. 77 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

Nesta tese propomos um modelo estocástico multimodal para regulação da expressão gênica em nível de transcrição. A definição de um espaço de parâmetros que contém o conteúdo biológico do sistema aliada à escolha apropriada de uma base para construir a matriz de acoplamento entre os estados do sistema levaram à obtenção de soluções exatas do modelo. Tais soluções são obtidas transformando as equações mestras em equações diferenciais parciais usando a técnica das funções geradoras e escrevendo os coeficientes das equações parciais em termos dos parâmetros biológicos do modelo. No regime estacionário obtivemos uma relação de recorrência para os coeficientes das séries de potências que definem as funções geradoras e a especificação das configurações de equilíbrio do sistema permite que estas séries sejam calculadas exatamente. Com as soluções exatas calculadas não só as distribuições de probabilidade foram obtidas como os momentos das distribuições. As distribuições de probabilidade de equilíbrio apresentam estruturas multimodais com vários picos e a análise do ruído (flutuação) mostra que a existência de um estado intermediário de eficiência transcricional leva a redução do ruído global do sistema. A inspeção dos autovalores da matriz de acoplamento mostrou que existem regiões onde a dinâmica dos momentos é de caráter oscilatória com amortecimento. Diferentes esquemas de acoplamento levam à diferentes regimes transientes, tal característica revela que o sistema multimodal apresentam maior flexibilidade adaptativa quando comparado com sistemas de um ou dois estados.

Palavras-chave: Expressão gênica. Ruído. Regulação transcricional. Processos estocásticos. Soluções exatas.

ABSTRACT

INNOCENTINI G.C.P. **A generalization of the spin–boson model for gene transcription under multiple control**. 2012. 77 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

In this thesis we propose a stochastic model for multimodal regulation of gene expression at the transcriptional level. The definition of a parameter space that contains the contents of the biological system coupled with the appropriate choice of a base to build the coupling matrix between the states of the system led to the exact solutions of the model. Such solutions are obtained by transforming master equations in partial differential equations using the technique of generating functions and writing the coefficients of partial equations in terms of biological parameters of the model. In the steady a recurrence relation for the coefficients of power series defining the generating functions was obtained and specification of the equilibrium configurations of the system allows the exact calculation of these series. With the exact solutions calculated not only the probability distributions were obtained but also the moments of the distributions. The equilibrium distributions probability is multimodal and presents several peaks. Analysis of the noise (fluctuation) shows that the existence of an state with intermediate transcriptional efficiency leads to a reduction of the overall system noise. Inspection of the eigenvalues of the coupling matrix showed that there are regions where the dynamics of the moments is damped oscillating. Different coupling schemes lead to different transient regimes, this feature reveals that the multimodal system have greater adaptive flexibility when compared to systems of one or two states.

Keywords: Gene expression. Noise. Transcriptional regulation. Stochastic process. Exact solutions

SUMÁRIO

1 Introdução **13**

2 O modelo para transcrição gênica e sua interpretação biológica **17**

2.1 O modelo para a transcrição 17

2.2 O espaço de parâmetros biológico 22

3 Propriedades assintóticas do processo transcricional **29**

3.1 As distribuições de probabilidade multimodais 31

3.2 Valor médio, flutuação e fator de Fano 35

3.3 Distribuição de probabilidade com valor médio fixo 43

4 Dinâmica dos momentos **47**

4.1 A solução dependente do tempo para os momentos das distribuições 47

4.2 Autovalores: regiões oscilantes 51

4.3 Comportamento dinâmico dos momentos para distintos esquemas de acoplamento 53

5 A generalização do modelo para um número arbitrário de estados **63**

5.1 O espaço de parâmentros biológicos generalizado 63

5.2 Distribuições de probabilidade para um sistema com quatro estados operacionais 65

6 Conclusões **69**

REFERÊNCIAS **72**

Capítulo 1

Introdução

Na década de trinta, do século passado, o contraste entre o crescente conhecimento sobre a matéria inanimada, alavancado pelos avanços na física, e o pouco entendimento sobre a matéria viva, fez com que fundações como o Institut de Biologie Physico-Chimique (IBPC) na França e a Rockefeller nos EUA , elaborassem e fortemente financiassem políticas para o desenvolvimento do estudo de sistemas biológicos tendo como prioridade a aplicação da metodologia e os conceitos oriundos da física e da química. Este marco onde estes três ramos da ciência são unidos, para resolver problemas que sozinhos não seriam capazes, é considerado o surgimento da biologia molecular (1).

A união de forças que resultou no nascimento da biologia molecular surtiu efeito já na década de quarenta e grandes descobertas foram alcançadas em um período de vinte e cinco anos (1940 - 1965) e entre elas estão: a relação de um gene uma enzima, a natureza dos genes, a estrutura de dupla hélice do DNA, o mecanismo de síntese proteica e o controle da expressão gênica. Essas cinco grandes descobertas foram cruciais para o melhor entendimento dos sistemas biológicos como também para o pleno desenvolvimento da biologia molecular porém, aqui, iremos nos ater à última

Os primeiros modelos para a regulação gênica foram elaborados a partir dos resultados obtidos ao se investigar dois sistemas: o estudo do ciclo lisogênico, um processo complexo de interação de uma bactéria e um fago e o estudo da adaptação de bactérias à novas fontes de nutrientes.

Na década de 50, André Lwoff e seus colegas, Elie Wollman e François Jacob, do

Instituto Pasteur observaram (2) que certa classe da bactéria *Escherichia coli* quando irradiada com doses moderadas de luz ultravioleta, paravam de crescer e aproximadamente 90 minutos mais tarde ocorria o rompimento das bactérias (lise) e uma quantidade de vírus, chamados de λ-phage, eram liberados no meio. Os vírus se multiplicavam infectando novas bactérias. Logo estas bactérias infectadas rompiam-se e novos vírus eram liberados. No entanto, algumas bactérias infectadas sobreviviam e carregavam o vírus em sua forma inativa. Essas bactérias cresciam e se dividiam normalmente até que a cultura era novamente irradiada e essas bactérias descendentes rompiam-se e liberavam uma nova quantidade de vírus no meio. André Lwoff e seus colegas chegaram à conclusão que esta alternância entre os dois estados do vírus – na forma inativa quando a bactéria está crescendo e se dividindo e a forma ativa quando a cultura é irradiada – é um simples exemplo do ligamento e desligamento dos genes.

Neste mesmo período, Monod observou, também em culturas da bactéria *Escherichia coli*, um fenômeno que ele batizou com o nome de diauxie. Neste, a cultura de bactéria é colocada para crescer em um meio composto por diversas fontes de açúcar, por exemplo, uma mistura de glicose e lactose. Monod constatou que a cultura crescia em dois estágios e que entre um estágio e outro havia um lapso de tempo em que a população de bactérias permanecia inalterada. Na primeira fase de crescimento a cultura crescia consumindo a glicose disponível no meio e na segunda consumindo lactose. Este comportamento populacional da *Escherichia coli* foi a inspiração para Monod propor no ano de 1961 (3,4), com a colaboração de Jacob, um modelo para o controle da expressão gênica. Neste modelo, Monod e Jacob, propõem que a enzima que degrada a lactose (β-galactosidase) é o produto da transcrição de um óperon (óperon *lac*). Este óperon é regulado, negativamente, por um gene constitutivo que produz o repressor *lacI*. Quando no meio em que a bactéria se desenvolve não está presente a lactose, o óperon *lac* encontra-se desligado com o repressor *lacI* em seu sítio operador. No entanto, quando a lactose está presente no meio, esta inativa o repressor e o óperon passa a ser expresso. Este processo regulatório ficou conhecido como o modelo de óperon onde, um gene regulador codifica uma proteína repressora que regula a transcrição de um conjunto de genes. Desde a elaboração do modelo de óperon um grande

progresso no entendimento e caracterização dos mecanismos envolvidos na regulação da expressão gênica (5) foi alcançado e mesmo assim duas hipóteses originalmente propostas por Monod e Jacob continuam válidas: a regulação de genes se dá principalmente em nível transcricional através da ação de proteínas que interagem diretamente com o DNA.

O desenvolvimento de técnicas eficientes para o sequenciamento de DNA ocorrido nas últimas décadas possibilitou a obtenção de uma quantidade volumosa de dados de genes de várias espécies, impulsionando de forma significativa a biologia celular e molecular. A resolução da estrutura terciária de proteínas a partir de métodos cristalográficos abriu o caminho para o entendimento da função biológica e da determinação de módulos operacionais. No entanto, a apropriação dessa quantidade imensa de informações não resultou no conhecimento da evolução celular na proporção esperada ou desejada. A cronologia e a geografia do processo de síntese proteica que demandam a previsão de quando e onde um gene se expressa estão longe de ser plenamente compreendidas. Para preencher esta lacuna, intensas investigações teóricas aliadas às experimentais (6–17) foram dedicadas para o entendimento mais amplo e profundo destas redes gênicas.

Estão envolvidas no processo de regulação da expressão gênica, uma grande quantidade de transformações químicas simultâneas. No entanto, o baixo número de moléculas de cada espécie presente na célula faz com que as flutuações moleculares sejam significativamente grandes (18) e uma descrição determinística, para este sistema biológico, deixa de ser a mais adequada fazendo com que uma modelagem estocástica seja inevitável (19–28). A partir dessa consideração, duas estratégias podem ser adotadas. A primeira consiste na introdução de flutuações, via o mecanismo de Langevin, com a inclusão de uma perturbação estocástica dependente do tempo (30–34), e a segunda, é a realização de simulações numéricas considerando o conjunto total de reações (35, 36) na célula.

Uma via alternativa às simulações numéricas de sistemas complexos e aos modelos que consistem de equações diferenciais ordinárias para as concentrações dos reagentes onde o ruído é incorporado via o mecanismo de Langevin é a de natureza microscópica onde o número de partículas produzidas por um único gene é a variável aleatória do modelo. Uma primeira classe de modelo microscópico foi proposta por van Oudenaarden (37,38). Neste

modelo o gene apresenta um único estado (uma única configuração da região controladora) e a transcrição é modelada por um processo de produção e degradação de mRNA a ser acoplado com um processo estocástico para a tradução. A introdução de mais de uma configuração para a região controladora foi proposta por Kepler (39) e mais tarde por Wolynes (40) onde o gene pode estar ligado (sem proteínas reguladoras obstruindo a transcrição) ou desligado (onde uma proteína reguladora se encontra no sítio operador obstruindo a transcrição) dando inicio à segunda classe de modelos microscópicos, a dos sistemas binários. As equações mestras que são utilizadas para descrever os sistemas binários foram resolvidas exatamente tanto em seu regime estacionário quanto dinâmico, para o caso em que o gene é autorregulado ou sofre regulação de um agente externo, (41–43) e as simetrias das equações desta classe de modelo foram identificadas (44, 45).

Nesta tese, inauguramos a terceira classe de modelos estocásticos microscópicos onde, primeiramente, consideramos um gene que pode operar em três estados de eficiência transcricional e, então, mostramos como esta classe de modelos pode ser generalizada para ser capaz de descrever um gene que pode produzir mRNA utilizando um número arbitrário de estados de eficiência transcricional.

Esta tese está organizada em capítulos onde: capítulo 2 apresentamos o modelo estocástico para um sistema de três estados e definimos o espaço de parâmetros biológicos; capítulo 3 exploramos as propriedades assintóticas do modelo; no capítulo 4 apresentamos o comportamento dinâmico das quantidades macroscópicas do modelo; capítulo 5 a generalização para um número arbitrário de estados e por fim, o último capítulo é dedicado às conclusões.

Capítulo 2

O modelo para transcrição gênica e sua interpretação biológica

Neste capítulo iremos apresentar as equações mestras do modelo que descrevem a transcrição de um gene sob múltiplo controle. Após discutir a forma e os parâmetros que compõem as equações passaremos a analisar a matriz de acoplamento do sistema e redefiniremos o espaço de parâmetros de forma a dar uma clara e precisa interpretação biológica ao mesmo.

2.1 O modelo para a transcrição

No capítulo anterior mencionamos os esforços teóricos, realizados nas últimas décadas, em se construir modelos que pudessem reproduzir e interpretar os dados experimentais disponíveis e com isso melhor compreender a expressão gênica. Apesar de avanços significativos terem sido alcançados ainda estão longe de serem suficientes para uma abrangente e mais completa compreensão dos mecanismos envolvidos na regulação de genes ou redes gênicas.

Destarte, neste trabalho propomos uma generalização do modelo de spins e bósons (40)

para a transcrição de um gene. Em sua formulação original o gene poderia apresentar dois estados funcionais: ligado ou desligado. Estando no estado ligado a transcrição ocorria sem obstruções e era total ou parcialmente suprimida quando o gene estava no estado desligado, constituindo assim um sistema de dois estados. A obtenção das soluções exatas, estacionárias e dependentes do tempo, do modelo binário (41–43) para a expressão gênica foi de grande importância na classificação e interpretação das propriedades biológicas do sistema identificando características intrínsecas e extrínsecas.

Apesar de ter sido bem sucedido, o modelo binário apresenta limitações uma vez que a existência de vários fatores de transcrição presentes na regulação gênica, atuando individualmente ou de forma combinada, faz com que o gene apresente inúmeros estados com diferentes níveis de eficiência transcricional. O paradigma destes sistemas multimoldais é o óperon *lac* responsável pelo transporte e metabolismo da lactose na *Escherichia coli* e outras bactérias (3, 4). Este óperon é constituído por uma região controladora, composto por um sítio promotor e um sítio operador, que regula, simultaneamente, três genes estruturais: *lacZ*, *lacY* e *lacA*. O produto do *lacZ* é a enzima β-galactosidase responsável pela digestão da lactose, o gene *lacY* codifica para a enzima permease que transporta a lactose para o meio intracelular e, por fim, o *lacA* codifica a transacetilase de β-galactosídeos, uma enzima que transfere um grupo acetil de acetil-CoA a β-galactosídeos. O controle do óperon *lac* é feito por regulação negativa através da proteína lac repressora codificada pelo gene constitutivo *lacI*. O mecanismo de regulação se dá da seguinte maneira: quando, no meio há uma alta concentração de glicose e baixa concentração de lactose o repressor *LacI* liga se ao sítio operador, da região controladora do óperon, obstruindo a transcrição do mesmo. No entanto, quando a lactose é presente em abundância e a glicose em baixas concentrações, a lactose, que entre na célula, liga se à proteína *lac* repressora fazendo com que o complexo lactose mais proteína *lac* repressora tenha uma afinidade muito menor com o sítio operador e, assim, não há associação da proteína reguladora com este sítio e a polimerase RNA encontra o caminho livre para realizar a transcrição do óperon. Um terceiro cenário emerge quando as concentrações de glicose e lactose são comparáveis. Em suma, temos três estados de eficiência transcricional. Estado reprimido (alta concen-

tração de glicose), estado ativado (alta concentração de lactose) e estado intermediário (as concentrações de lactose e glicose são comparáveis).

Com este exemplo em mente, propomos um modelo estocástico onde a variável aleatória, n, é o número de mRNA presente na célula e escrevemos equações mestras acopladas, do tipo nascimento e morte, que governam a dinâmica das distribuições de probabilidade de encontrarmos o óperon em um dos seus três estados de eficiência transcricional, no instante de tempo t e com n moléculas de mRNA presentes. As equações para as distribuições são:

$$\begin{aligned}
\frac{d\phi_n^1}{dt} &= k_1[\phi_{n-1}^1 - \phi_n^1] + \rho[(n+1)\phi_{n+1}^1 - n\phi_n^1] - h_1\phi_n^1 + h_2\epsilon_2\phi_n^2 + h_3(1-\epsilon_3)\phi_n^3, \quad (2.1)\\
\frac{d\phi_n^2}{dt} &= k_2[\phi_{n-1}^2 - \phi_n^2] + \rho[(n+1)\phi_{n+1}^2 - n\phi_n^2] + h_1(1-\epsilon_1)\phi_n^1 - h_2\phi_n^2 + h_3\epsilon_3\phi_n^3,\\
\frac{d\phi_n^3}{dt} &= k_3[\phi_{n-1}^3 - \phi_n^3] + \rho[(n+1)\phi_{n+1}^3 - n\phi_n^3] + h_1\epsilon_1\phi_n^1 + h_2(1-\epsilon_2)\phi_n^2 - h_3\phi_n^3.
\end{aligned}$$

Onde $\phi_n^1(t)$, $\phi_n^2(t)$ e $\phi_n^3(t)$, são as distribuições de probabilidade de encontrarmos o óperon no estado ativado, intermediário e reprimido, respectivamente. Os parâmetros k_1, k_2 e k_3 são as taxas de produção de mRNA de cada um dos estados e ρ a taxa de decaimento das moléculas de mRNA, que é a mesma para os três estados uma vez que o tempo de vida médio da molécula de mRNA, na célula, não depende do estado do sistema em que foi produzida. Os parâmetros restantes são as taxas de transições entre os estados do sistema e podem ser representadas pelo diagrama:

$$\phi_n^1 \underset{h_2\epsilon_2}{\overset{h_1(1-\epsilon_1)}{\rightleftharpoons}} \phi_n^2, \qquad \phi_n^1 \underset{h_3(1-\epsilon_3)}{\overset{h_1\epsilon_1}{\rightleftharpoons}} \phi_n^3, \qquad \phi_n^2 \underset{h_3\epsilon_3}{\overset{h_2(1-\epsilon_2)}{\rightleftharpoons}} \phi_n^3.$$

com $0 \leq h_i < \infty$, para $i = 1, 2$ e 3, e $0 \leq \epsilon_i \leq 1$, também com $i = 1, 2$ e 3.

As equações mestras em 2.1 são equações lineares em n e do tipo diferencial-diferença. Usando a técnica das funções geradoras (46, 47) podemos transformá-las em equações à derivadas parciais e com isso lidar com objetos mais comumente encontrado em problemas de física com técnicas para resolvê-los que são inerentes à física matemática. Para tanto, propomos

$$\phi^1(z,t) = \sum_{n=0}^{\infty} \phi_n^1(t) z^n, \qquad \phi^2(z,t) = \sum_{n=0}^{\infty} \phi_n^2(t) z^n, \qquad \phi^3(z,t) = \sum_{n=0}^{\infty} \phi_n^3(t) z^n, \qquad (2.2)$$

onde z é uma variável complexa que pertence ao disco unitário, $z \in [0,1]$. As equações à derivadas parciais resultantes, nas variáveis t e z, são:

$$\begin{aligned}
\frac{\partial \phi^1}{\partial t} &= (z-1)[k_1\phi^1 - \rho\frac{\partial \phi^1}{\partial z}] - h_1\phi^1 + h_2\epsilon_2\phi^2 + h_3(1-\epsilon_3)\phi^3, \\
\frac{\partial \phi^2}{\partial t} &= (z-1)[k_2\phi^2 - \rho\frac{\partial \phi^2}{\partial z}] + h_1(1-\epsilon_1)\phi^1 - h_2\phi^2 + h_3\epsilon_3\phi^3, \\
\frac{\partial \phi^3}{\partial t} &= (z-1)[k_3\phi^3 - \rho\frac{\partial \phi^3}{\partial z}] + h_1\epsilon_1\phi^1 + h_2(1-\epsilon_2)\phi^2 - h_3\phi^3.
\end{aligned} \tag{2.3}$$

As distribuições de probabilidade são reobtidas através de derivadas das funções geradoras quando avaliadas em $z = 0$, da seguinte maneira:

$$\begin{aligned}
\phi_n^1(t) &= \left[\frac{1}{n!}\frac{\partial^n \phi^1(z,t)}{\partial z^n}\right]_{z=0}, \\
\phi_n^2(t) &= \left[\frac{1}{n!}\frac{\partial^n \phi^2(z,t)}{\partial z^n}\right]_{z=0}, \\
\phi_n^3(t) &= \left[\frac{1}{n!}\frac{\partial^n \phi^3(z,t)}{\partial z^n}\right]_{z=0}.
\end{aligned} \tag{2.4}$$

Os momentos das distribuições de probabilidade podem, igualmente, ser calculados através de derivadas das funções geradoras mas, aqui, avaliadas em $z = 1$, através das fórmulas:

$$\begin{aligned}
\left\langle n_1^j \right\rangle (t) &= \left[\left(z\frac{\partial}{\partial z}\right)^j \phi^1(z,t)\right]_{z=1} = \sum_{n=0}^{\infty} n^j \phi_n^1(t), \\
\left\langle n_2^j \right\rangle (t) &= \left[\left(z\frac{\partial}{\partial z}\right)^j \phi^2(z,t)\right]_{z=1} = \sum_{n=0}^{\infty} n^j \phi_n^2(t), \\
\left\langle n_3^j \right\rangle (t) &= \left[\left(z\frac{\partial}{\partial z}\right)^j \phi^3(z,t)\right]_{z=1} = \sum_{n=0}^{\infty} n^j \phi_n^3(t),
\end{aligned} \tag{2.5}$$

ao contrário das equações anteriores, aqui, nas fórmulas para os momentos das distribuições, o índice superior, j, indentifica a ordem do momento e o índice inferior, $i = 1, 2$ e 3, o estado considerado.

Antes de passarmos a investigar o espaço de parâmetros em termos de quantidades biológicas (que será feito na próxima seção) iremos reescrever as equações do modelo, mas desta vez, usando a notação vetorial que, como será apresentado no penúltimo capítulo desta tese, é de grande utilidade para a generalização do modelo onde um número arbitrário de estados pode ser considerado. Na notação de vetores e matrizes, temos para as

equações mestras:

$$\frac{d\vec{\phi}_n(t)}{dt} = K[\vec{\phi}_{n-1}(t) - \vec{\phi}_n(t)] + \rho[(n+1)\vec{\phi}_{n+1}(t) - n\vec{\phi}_n(t)] + H\vec{\phi}_n(t), \tag{2.6}$$

onde $\vec{\phi}_n(t) = (\phi_n^1(t), ..., \phi_n^N(t))^T$, e N é o número de estados do sistema. As matrizes K e ρ são ambas diagonais, de ordem $N \times N$, sendo que os elementos diagonais da primeira são as diferentes taxas de produção de mRNA, que correspondem a cada um dos estados de eficiência transcricional do sistema, e a última proporcional à identidade. A matriz H é a matriz de acoplamento do sistema e tem as seguintes propriedades:

$$\sum_{i=0}^{N} H_{ij} = 0, \quad H_{ii} \leq 0. \tag{2.7}$$

No caso específico que estamos tratando com $N = 3$, a matriz de acoplamento tem a forma:

$$H = \begin{pmatrix} -h_1 & h_2\epsilon_2 & h_3(1-\epsilon_3) \\ h_1(1-\epsilon_1) & -h_2 & h_3\epsilon_3 \\ h_1\epsilon_1 & h_2(1-\epsilon_2) & -h_3 \end{pmatrix} \tag{2.8}$$

As equações mestras na forma vetorial podem ser reescritas em termos de equações à derivadas parciais, utilizando se do vetor

$$\vec{\phi}(z,t) = \sum_{n=0}^{\infty} \vec{\phi}_n(t) z^n, \tag{2.9}$$

o que leva ao sistema de equações diferenciais parciais acopladas

$$\frac{\partial\vec{\phi}(z,t)}{\partial t} = (z-1)[K\vec{\phi}(z,t) - \rho\hat{1}\frac{\partial\vec{\phi}(z,t)}{\partial z}] + H\vec{\phi}(z,t). \tag{2.10}$$

O vetor cujo as componentes são as distribuições de probabilidade de cada estado é reobtido através de derivadas usando:

$$\vec{\phi}_n(t) = \left[\frac{1}{n!}\frac{\partial^n\vec{\phi}(z,t)}{\partial z^n}\right]_{z=0}, \tag{2.11}$$

e os momentos da seguinte maneira:

$$\left\langle \vec{n}^j \right\rangle = \left[\left(z\frac{\partial}{\partial z}\right)^j \vec{\phi}(z,t)\right]_{z=1}. \tag{2.12}$$

onde j é a ordem do momento considerado.

2.2 O espaço de parâmetros biológico

As equações mestras do nosso modelo podem ser interpretadas como uma combinação de dois processos estocásticos (46, 47). No primeiro, temos a produção-degradação de moléculas de mRNA, onde a produção é controlada pelas taxas k_1, k_2 e k_3 e a degradação pela taxa ρ, que é a mesma para todos os estados. O segundo é um processo telegráfico que leva em consideração todas as possíveis transições entre os estados do sistema.

A primeira redefinição do espaço de parâmetros envolve as taxas que governam o processo de produção-degradação de mRNA. Ao invés de lidarmos com as taxas, iremos considerar o parâmetro de eficiência transcricional de cada um dos estados. Este parâmetro é obtido pela razão entre cada taxa de produção pela de degradação, dando:

$$n_1 = \frac{k_1}{\rho}, \quad n_2 = \frac{k_2}{\rho}, \quad n_3 = \frac{k_3}{\rho}. \tag{2.13}$$

Quanto maior for n_i, em relação aos demais, mais eficiente na produção de mRNA será o estado i, ou seja, um estado eficiente produzirá, em média, dezenas de cópias de mRNA, ao passo que, um estado de baixa eficiência produzirá algumas unidades ou não produzirá mRNA.

O processo telegráfico é controlado pelas taxas de transição entre os estados do sistema que são independentes da variável aleatória, n. No entanto, estas taxas dependem tanto de propriedades intrínsecas do gene, como a afinidade do operador por uma proteína reguladora específica, quanto de propriedades que são extrínsecas, dependendo das condições do meio, como por exemplo, a disponibilidade de proteínas reguladoras. Com o intuito de desacoplar a dependência intrínseca daquela proveniente do meio em que o sistema está, consideramos, apenas, a parte do processo telegráfico do modelo, que pode ser obtida tomando a média das equações mestras em 2.1 com relação ao número de moléculas de mRNA, n, ou, equivalentemente, colocando $z = 1$ nas Eqs.(2.3), resultando

nas equações:

$$\begin{aligned}
\frac{dP_1(t)}{dt} &= -h_1P_1(t) + h_2\epsilon_2P_2(t) + h_3(1-\epsilon_3)P_3(t) \\
\frac{dP_2(t)}{dt} &= h_1(1-\epsilon_1)P_1(t) - h_2P_2(t) + h_3\epsilon_3P_3(t) \\
\frac{dP_3(t)}{dt} &= h_1\epsilon_1P_1(t) + h_2(1-\epsilon_2)P_2(t) - h_3P_3(t)
\end{aligned} \tag{2.14}$$

onde as funções dependentes do tempo $P_i(t)$ são $\phi^i(z=1,t)$ (com i=1,2 e 3) e representam as probabilidades de ocupação de se encontrar o gene em um de seus modos operacionais independentemente do número de mRNA, n. Alternativamente, podemos ter as Eqs.(2.14), em sua forma vetorial fazendo $z=1$ nas Eqs.(2.10) e teremos:

$$\frac{d\vec{P}(t)}{dt} = H\vec{P}(t), \tag{2.15}$$

onde $\vec{P}(t)$=$(P_1(t), P_2(t), P_3(t))^T$ e H é a matriz de acoplamento assim como dada na Eq.(2.8). A solução do sistema de equações em 2.15 será apresentada no capítulo 4 desta tese, juntamente com a solução dependente do tempo dos primeiros e segundos momentos das distribuições. Aqui iremos explorar as propriedades da matriz de acoplamento, H, e reinterpretar seus parâmetros em termos de quantidades biológicas do sistema. Uma primeira característica desta matriz é que ela pode ser fatorada como o produto de outras duas:

$$H = \epsilon h \tag{2.16}$$

onde a matriz

$$\epsilon = \begin{pmatrix} -1 & \epsilon_2 & (1-\epsilon_3) \\ (1-\epsilon_1) & -1 & \epsilon_3 \\ \epsilon_1 & (1-\epsilon_2) & -1 \end{pmatrix} \tag{2.17}$$

e a matriz

$$h = \begin{pmatrix} h_1 & 0 & 0 \\ 0 & h_2 & 0 \\ 0 & 0 & h_3 \end{pmatrix}. \tag{2.18}$$

Inspeção da matriz H em sua forma fatorada mostra que a matriz ϵ é responsável pela existência do acoplamento entre os estados do sistema. A seleção dos valores para

os elementos da matriz ϵ irá definir se um estado está conectado com outros ou não e, também, o quão intenso é este acoplamento. Por exemplo, se $\epsilon_1 = 1$, o estado 1 estará fortemente acoplado ao estado 3 e não estará influenciando o estado 2, ao passo que $\epsilon_2 = 1$ faz com que o acoplamento entre os estados 1 e 2 seja forte e entre os estados 2 e 3 não haja comunicação, etc. No limite em que todos os $\epsilon_i = 0$ nós temos um modelo onde o acoplamento entre os estados é cíclico ($1 \rightharpoonup 2 \rightharpoonup 3 \rightharpoonup 1$), quando todos os $\epsilon_i = 1$ estamos no regime de acoplamento anticíclico ($1 \rightharpoonup 3 \rightharpoonup 2 \rightharpoonup 1$) e para os $\epsilon_i = 1/2$ todos os estados estão acoplados com os demais e com mesma intensidade. A forma e a intensidade do acoplamento depende do sistema em consideração e, assim, vemos que os elementos da matriz ϵ são propriedades intrínsecas do sistema. A matriz h carrega a informação sobre a ocupação de um determinado estado. Para fazer esta afirmação mais clara, vamos escrever os elementos da matriz h em termos das probabilidades de ocupação assintóticas de encontrarmos o sistema em um de seus estados. Para tanto, propomos

$$\vec{P}(t) = \exp(\lambda t)\vec{q} \tag{2.19}$$

onde λ é uma matriz proporcional à identidade. Substituindo na Eq.(2.15), temos a equação:

$$H\vec{q} = \lambda\vec{q}. \tag{2.20}$$

Este é agora um problema de autovalores e autovetores do operador H. As informações sobre o estado assintótico estão codificadas no autovetor correspondente ao autovalor nulo. As propriedades de uma matriz do tipo H, expressas nas Eqs.(2.7), garantem que o determinante de tal matriz sempre é igual à zero. Como o produto dos autovalores de uma matriz é idêntico ao determinante da mesma, ao menos um autovalor nulo existe. Assim sendo, calculando o autovetor correspondente ao autovalor nulo da Eq.(2.20), obtemos:

$$\vec{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} h_3 h_2 \delta_1 \\ h_1 h_3 \delta_2 \\ h_2 h_1 \delta_3 \end{pmatrix}, \tag{2.21}$$

onde os δ_i são os cofatores dos elementos da diagonal da matriz ϵ ($\delta_i = C_{ii}(\epsilon)$), para

$i = 1, 2, 3$) e têm as expressões:

$$\begin{aligned} \delta_1 &= 1 - \epsilon_3 + \epsilon_3\epsilon_2 \\ \delta_2 &= 1 - \epsilon_1 + \epsilon_1\epsilon_3 \\ \delta_3 &= 1 - \epsilon_2 + \epsilon_2\epsilon_1. \end{aligned} \tag{2.22}$$

O autovetor, $\vec{q}$, apresenta a seguinte estrutura: sua componente i é o produto dos cofatores dos elementos da diagonal das matrizes ϵ e h, ou seja $q_i = C_{ii}(\epsilon)C_{ii}(h)$. Propriamente normalizado, este autovetor é o vetor de probabilidade de ocupação assintótica do sistema que contêm a informação sobre a configuração operacional do mesmo. O vetor de probabilidade de ocupação é

$$\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} q_1/q \\ q_2/q \\ q_3/q \end{pmatrix} \tag{2.23}$$

onde o escalar q é o produto escalar do vetor $\vec{q}$ com o autovetor à esquerda de H, $\vec{u} = (1, 1, 1)$, correspondente ao autovalor nulo ($\vec{u}H = 0$), assim, temos

$$q = \vec{u}.\vec{q} = q_1 + q_2 + q_3 = h_3h_2\delta_1 + h_1h_3\delta_2 + h_2h_1\delta_3. \tag{2.24}$$

A matriz h tem três parâmetros livres e as equações em 2.23 formam um sistema linear homogêneo onde apenas duas equações são linearmente independentes pelo fato que $p_1 + p_2 + p_3 = 1$. Para expressarmos os elementos da matriz h em termos das quantidades biológicas do sistema precisamos de mais uma equação que envolva os h_i. Esta equação é obtida considerando se o traço da matriz H,

$$-\operatorname{Tr}(H) = \Sigma = h_1 + h_2 + h_3 \tag{2.25}$$

levando em conta, junto com a Eq.(2.25), as equações,

$$\frac{q_1}{q} = p_1 \qquad \text{e} \qquad \frac{q_2}{q} = p_2, \tag{2.26}$$

nós obtemos os parâmetros da matriz h em termos das quantidades biológicas do modelo

como:

$$\begin{aligned} h_1 &= \Sigma\delta_1 p_2 p_3/\Delta \\ h_2 &= \Sigma p_1 \delta_2 p_3/\Delta \\ h_3 &= \Sigma p_1 p_2 \delta_3/\Delta \end{aligned} \tag{2.27}$$

onde a quantidade

$$\Delta = \delta_1 p_2 p_3 + p_1 \delta_2 p_3 + p_1 p_2 \delta_3. \tag{2.28}$$

Os parâmetros p_1, p_2 e p_3, são as probabilidades de ocupação assintótica de cada estado. Elas dependem das condições do meio em que o sistema se encontra e portanto, são propriedades extrínsecas. No caso do óperon-lac estes parâmetros seriam funções das concentrações de glicose e lactose presentes no meio celular. Para discutir o significado do parâmetro Σ apresentamos os autovalores não nulos da matriz H,

$$\lambda^{\pm} = -\frac{\Sigma}{2}(1 \pm \sqrt{1 - 4q/\Sigma^2}), \tag{2.29}$$

ou, colocando a expressão de q, explicitamente em função das quantidades biológicas do modelo,

$$q = \frac{\Sigma^2 \delta_1 \delta_2 \delta_3 p_1 p_2 p_3}{\Delta^2}, \tag{2.30}$$

temos

$$\lambda^{\pm} = -\frac{\Sigma}{2}\left(1 \pm \sqrt{1 - \frac{4\delta_1\delta_2\delta_3 p_1 p_2 p_3}{\Delta^2}}\right). \tag{2.31}$$

A análise dos autovalores não nulos da matriz H mostra que Σ é a componente principal que governa a dinâmica do sistema até este estar em sua configuração assintótica. O modelo tem duas escalas de tempo importante: $1/\rho$, o tempo de vida médio de uma molécula de mRNA e $1/\Sigma$, o termo dominante de decaimento do sistema. A razão Σ/ρ mede a flexibilidade do switch. Para grandes valores desta razão dizemos que o switch é rápido ao passo que, para valores pequenos estamos na região de switch lento. Os adjetivos rápido e lento utilizados para qualificar o switch são para serem interpretados como o quão hábil é o switch para alcançar a configuração de equilíbrio e essa é uma propriedade intrínseca do sistema. Neste espaço de parâmetros biológico, a matriz H

assume a forma:

$$H = \frac{\Sigma}{\Delta}\begin{pmatrix} -1 & \epsilon_2 & (1-\epsilon_3) \\ (1-\epsilon_1) & -1 & \epsilon_3 \\ \epsilon_1 & (1-\epsilon_2) & -1 \end{pmatrix}\begin{pmatrix} \delta_1 p_2 p_3 & 0 & 0 \\ 0 & p_1\delta_2 p_3 & 0 \\ 0 & 0 & p_1 p_2 \delta_3 \end{pmatrix}. \tag{2.32}$$

Antes de passarmos para os próximos capítulos, por uma questão de completeza, vamos apresentar os outros dois autovetores que correspondem a $\lambda^{\pm}$. Para tanto, notamos que:

$$\vec{u}H\vec{v}^{\pm} = \lambda^{\pm}\vec{u}.\vec{v}^{\pm} = 0. \tag{2.33}$$

Então, a estrutura para $\vec{v}^{\pm}$ fica sendo

$$\vec{v}^{\pm} = \begin{pmatrix} 1 \\ A^{\pm} \\ -1 - A^{\pm} \end{pmatrix} \tag{2.34}$$

onde temos

$$A^{\pm} = \frac{p_1 p_2 \delta_3 (1-\epsilon_3) + \delta_1 p_2 p_3 + \Delta\lambda^{\pm}/\Sigma}{p_1 \delta_2 p_3 \epsilon_2 - (1-\epsilon_3) p_1 p_2 \delta_3} \tag{2.35}$$

Iremos utilizar este resultado no Capítulo 4 desta tese onde iremos mostrar o comportamento dinâmico dos momentos das distribuições. No que segue, Capítulo 3, iremos apresentar as distribuições de probabilidade estacionárias e analisar o ruído do sistema uma vez que este está em sua configuração de equilíbrio. Todas as fórmulas que serão apresentadas estarão expressas em termos dos parâmetros biológico aqui definidos e suas implicações e interpretações serão exploradas.

Capítulo 3

Propriedades assintóticas do processo transcricional

Neste capítulo, focamos nossa atenção nas soluções estacionárias do processo de transcrição e suas características fundamentais. Iremos apresentar as fórmulas para se calcular as distribuições de probabilidade assim como as fórmulas para o valor médio e o ruído do sistema.

As propriedades assintóticas são obtidas a partir da versão estacionária da Eq.(2.10),

$$(z-1)\frac{d\vec{\phi}(z)}{dz} = (z-1)N\vec{\phi}(z) + \frac{1}{\rho}H\vec{\phi}(z). \tag{3.1}$$

onde $\vec{\phi}(z) = (\phi^1(z), \phi^2(z), \phi^3(z))^T$, H é a matriz de acoplamento assim como em 2.8 e a matriz

$$N = \begin{pmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{pmatrix}, \tag{3.2}$$

é a matriz de eficiência transcricional. Observamos que toda a equação 2.10 foi multiplicada pela matriz ρ^{-1} para ter a forma estacionária assim como na Eq.(3.1). Ao invés de utilizar a variável z propomos a substituição $w = z - 1$ e obtemos a EDO,

$$w\frac{d\vec{\phi}(w)}{dw} = wN\vec{\phi}(w) + \frac{1}{\rho}H\vec{\phi}(w). \tag{3.3}$$

Introduzindo a série

$$\vec{\phi}(w) = \sum_{n=0}^{\infty} \vec{\phi}_n w^n \tag{3.4}$$

substituindo na Eq.(3.3) e colecionando termos em potências de w^n obtemos a seguinte relação de recorrência:

$$\vec{\phi}_n = \frac{1}{n - H/\rho} N \vec{\phi}_{n-1}. \tag{3.5}$$

Primeiro, gostaríamos de salientar a importância na mudança de variáveis, $z \to w$. Aplicando a Eq.(2.12) na expansão em potência de w proposta na Eq.(3.4), podemos identificar os primeiro termos como:

$$\vec{\phi}(w) = \vec{p} + \langle \vec{n} \rangle \, w + \frac{\langle \vec{n}^2 \rangle - \langle \vec{n} \rangle}{2} w^2 + \mathrm{O}(w^3). \tag{3.6}$$

Onde $\vec{p}$ é o vetor cuja as componentes são as probabilidades de ocupação de cada estado, $\langle \vec{n} \rangle$ e o vetor valor médio, e assim por diante. Ou seja, a expansão em potências da variável w pode ser interpretada como a expansão em momentos da distribuição. Então, nesta variável e com os parâmetros como funções das quantidades biológicas do modelo, ao colocar $n = 1$ na Eq.(3.5), vemos que $\vec{\phi}_1 \propto \vec{\phi}_0 = \vec{p}$ e, assim, para se calcular todos os coeficientes da série de potências em w basta especificar o vetor $\vec{\phi}_0 = \vec{p}$.

A matriz $U = (n - H/\rho)$ pode ser colocada em uma forma muita mais conveniente usando o fato que qualquer matriz é solução de seu polinômio característico, assim, temos para a matriz U

$$U^3 + \mathrm{Trace}(U)U^2 + (C_{11}(U) + C_{22}(U) + C_{33}(U))U + \det(U) = 0, \tag{3.7}$$

onde os $C_{ii}(U)$ são os cofatores dos elementos da diagonal da matriz U. Resolvendo a Eq.(3.7) para U obtemos

$$\frac{1}{U} = -\frac{U^2 + \mathrm{Trace}(U)U + (C_{11}(U) + C_{22}(U) + C_{33}(U))}{\det(U)}. \tag{3.8}$$

Como $U = (n - H/\rho)$ o termo em U^2 irá apresentar um termo em H^2 no entanto, temos

$$H^2 = H\Sigma - q(1 - P), \tag{3.9}$$

com 1 sendo a matriz identidade e a matriz

$$P = \begin{pmatrix} p_1 & p_1 & p_1 \\ p_2 & p_2 & p_2 \\ p_3 & p_3 & p_3 \end{pmatrix}. \tag{3.10}$$

Desta forma, substituindo $U = (n - H/\rho)$ na Eq.(3.8), colecionando termos em ordem de H e trocando H^2 por sua expressão, assim como dada na Eq.(3.9), temos

$$\frac{1}{n - H/\rho} = \frac{n\rho(n\rho + \Sigma)1 + qP + n\rho H}{n(n^2\rho^2 + n\rho\Sigma + q)}. \tag{3.11}$$

Utilizando o resultado 3.11 e substituindo na Eq.(3.5) a relação de recorrência para os coeficientes da expansão em w fica da forma

$$\vec{\phi}_n = \frac{n\rho(n\rho + \Sigma)1 + qP + n\rho H}{n(n^2\rho^2 + n\rho\Sigma + q)} N\vec{\phi}_{n-1}. \tag{3.12}$$

Especificando se os parâmetros de eficiência do gene (elementos da matriz N), os acoplamentos entre os estados (elementos da matriz ϵ) e as probabilidades de ocupação assintóticas, podemos utilizar a relação de recorrência em 3.12 para calcular as distribuições de probabilidade exatas do sistema e, também, todos os momentos das distribuições. Para se obter as distribuições de probabilidade, uma vez calculado os coeficientes da expansão em w, pode se adotar duas estratégias: uma é trocar $w = z - 1$ no polinômio resultante, usar a expansão binomial para se calcular $(z - 1)^n$ e colecionar potências de z onde os coeficientes de z^n serão vetores cuja as componentes serão as probabilidades de encontramos o sistema em um de seus estados com n moléculas de mRNA na célula ou, de forma análoga, podemos usar a Eq.(2.11), tomando o cuidado de substituir z por w e não mais avaliando em $z = 0$ mas em $w = -1$. Os momentos são obtidos diretamente da expansão em w, como ressaltamos ao comentar a Eq.(3.6).

3.1 As distribuições de probabilidade multimodais

Nesta seção apresentamos as distribuições de probabilidade calculadas usando se a Eq.(3.12) e a estratégia da expansão binomial uma vez que a variável w é trocada por $z-1$ no polinômio resultante. Para todos os gráfico que apresentamos nesta seção os parâmetros de eficiência transcricional são os mesmo respeitando a hierarquia $n_1 > n_2 > n_3$ e com valores: $n_1 = 40$, $n_2 = 20$ e $n_3 = 5$. Os elementos da matriz ϵ foram todos colocados iguais a $1/2$. Neste esquema de acoplamento, todos os estados estão conectados e com mesma intensidade. Outros esquemas de acoplamento serão apresentados no Capítulo 4 desta tese onde iremos exibir o comportamento dinâmico dos momentos das distribuições e sete esquemas distintos de acoplamento foram escolhidos e detalhados.

As distribuições de probabilidade desta seção foram calculadas no caso em que o switch é lento ($\Sigma = 0.5$) e no caso em que o switch é rápido ($\Sigma = 5$). Para cada regime de flexibilidade do gene nós usamos quatro configurações assintóticas distintas. A primeira é quando, no equilíbrio, o sistema pode estar, com mesma probabilidade, em qualquer um de seus estados para isso $\vec{p} = (1/3, 1/3, 1/3)^T$. Os outros regimes são caracterizados pela maior probabilidade de encontrar o sistema em um estado específico. Este opera predominantemente no estado i quando p_i é muito maior que os demais. Para acomodar esta situação escolhemos: predominância do estado de maior eficiência transcricional, $\vec{p} = (0.8, 0.1, 0.1)^T$, predominância do estado intermediário, $\vec{p} = (0.1, 0.8, 0.1)^T$ e por fim, predominância do estado de eficiência basal, $\vec{p} = (0.1, 0.1, 0.8)^T$.

Na Fig.(3.1) exibimos as distribuições de probabilidade de cada modo $\phi_n^1, \phi_n^2, \phi_n^3$ e a total $\phi_n = \phi_n^1 + \phi_n^2 + \phi_n^3$, para o caso em que o switch é lento. No primeiro painel, Fig.3.1(a), as probabilidades de ocupação assintóticas são iguais ($p_1 = 1/3$, $p_2 = 1/3$ e $p_3 = 1/3$). Cada modo operacional forma um pico resultando na probabilidade total de três picos indicando que o sistema opera em todos os seus estados para produzir mRNA. Na Fig.3.1(b) o valor das probabilidades de ocupação são: $p_1 = 0.8$, $p_2 = 0.1$ e $p_3 = 0.1$. Nota se um pico bastante pronunciado centrado em torno de $n = 40$ e dois picos residuais para $n \approx 20$ e $n \approx 5$ indicando que o sistema produz mRNA predominantemente em seu modo de máxima eficiência. No painel 3.1(c) temos $p_1 = 0.1$, $p_2 = 0.8$ e $p_3 = 0.1$ e o sistema é mais provável de ser encontrado produzindo no estado de eficiência inter-

mediária como pode ser visto pela existência do pico ao redor de $n = 20$. Na Fig.3.1(d) o cenário é o de se encontrar o sistema sendo transcrito estando em seu estado de menor eficiência e as ocupâncias têm os valores: $p_1 = 0.1$, $p_2 = 0.1$ e $p_3 = 0.8$. A modificação no valor das probabilidades assintóticas de ocupação de cada estado que acontece de um painel para outro na Fig.(3.1) leva a distribuição que tem uma estrutura de multipicos à estruturas quase poissonianas. No caso em que a distribuição é de multipicos os momentos da distribuição, valor médio e variância, comumente usados para descrever sistemas estocásticos em algum esquema de aproximação, mostram se ineficazes para capturar esta característica.

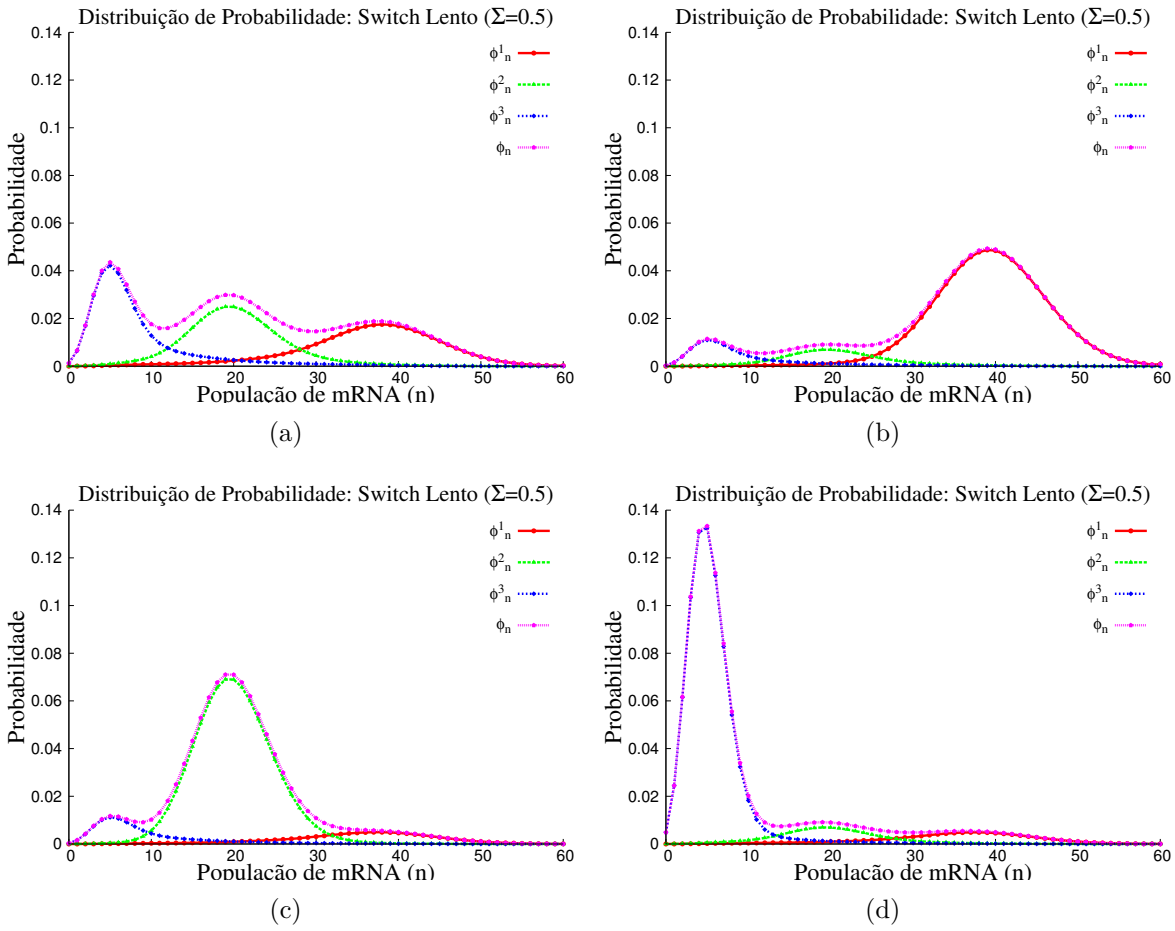

Figura 3.1 – Neste conjunto de gráficos mostramos o comportamento das distribuições de probabilidade de cada modo de eficiência transcricional assim como para a probabilidade total ($\phi_n = \phi_n^1 + \phi_n^2 + \phi_n^3$) no regime de switch lento, $\Sigma = 0.5$, calculadas usando a Eq.(3.12). Na Fig.3.1(a) temos $p_1 = 1/3$, $p_2 = 1/3$ e $p_3 = 1/3$, Fig.3.1(b) temos $p_1 = 0.8$, $p_2 = 0.1$ e $p_3 = 0.1$, Fig.3.1(c) temos $p_1 = 0.1$, $p_2 = 0.8$ e $p_3 = 0.1$ e Fig.3.1(d) temos $p_1 = 0.1$, $p_2 = 0.1$ e $p_3 = 0.8$. Para todos os gráficos os valores dos demais parâmetros do modelo são: $n_1 = 40$, $n_2 = 20$ e $n_3 = 5$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1/2$ e $\rho = 1$.

Na Fig.(3.2) mostramos as distribuições de probabilidade de cada modo e a total para o mesmo esquema de acoplamento e com os mesmos valores das probabilidades de ocupação

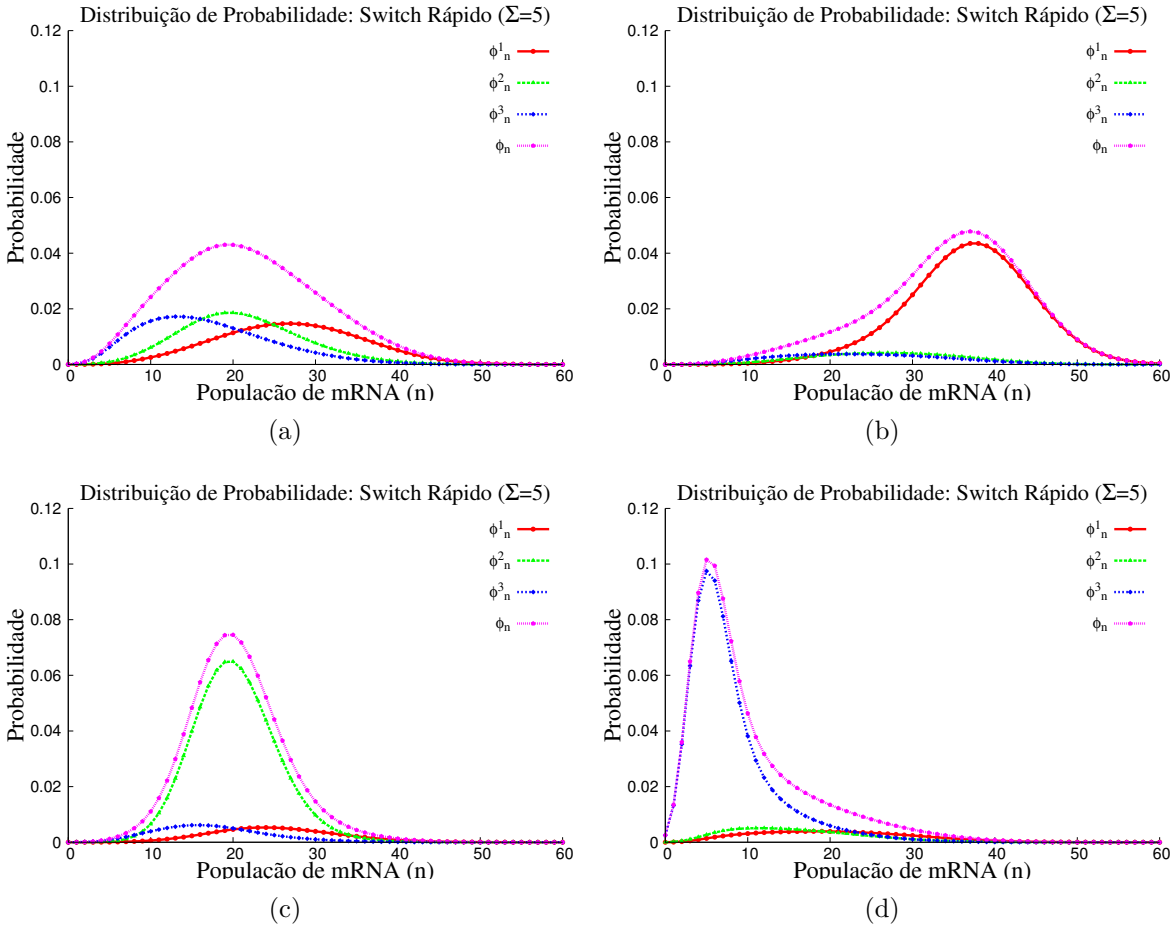

Figura 3.2 – Neste conjunto de gráficos mostramos o comportamento das distribuições de probabilidade de cada modo de eficiência transcricional assim como para a probabilidade total ($\phi_n = \phi_n^1 + \phi_n^2 + \phi_n^3$) no regime de switch lento, $\Sigma = 5$, calculadas usando a Eq.(3.12). Na Fig.3.2(a) temos $p_1 = 1/3$, $p_2 = 1/3$ e $p_3 = 1/3$, Fig.3.2(b) temos $p_1 = 0.8$, $p_2 = 0.1$ e $p_3 = 0.1$, Fig.3.2(c) temos $p_1 = 0.1$, $p_2 = 0.8$ e $p_3 = 0.1$ e Fig.3.2(d) temos $p_1 = 0.1$, $p_2 = 0.1$ e $p_3 = 0.8$. Para todos os gráficos os valores dos demais parâmetros do modelo são: $n_1 = 40$, $n_2 = 20$ e $n_3 = 5$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1/2$ e $\rho = 1$.

que usamos na figura anterior. No entanto, neste conjunto de gráficos o gene é rápido, $\Sigma = 5$. Comparando os painéis da Fig.(3.1) com os da Fig.(3.2) três efeitos principais podem ser detectados quando se aumenta a habilidade do gene de alcançar a configuração de equilíbrio: a distribuição de três picos desaparece totalmente, os picos individuais de cada modo decrescem e estes são deslocados em comparação com o caso de switch lento. Na Fig.3.2(a) podemos ver que a distribuição ϕ_n^1 é centrada em $n \approx 28$ e ϕ_n^3 centrada em $n \approx 12$ o primeiro deslocou se de $n \approx 40$ e o segundo de $n \approx 5$. Os próximos painéis exibem as distribuições de probabilidade quando um dos estados é dominante, começando com o estado 1 sendo o mais provável ($p_1 = 0.8$) na Fig.3.2(b) e decrescendo até chegar na Fig.3.2(d) onde a dominância é do estado 3 ($p_3 = 0.8$).

Para valores extremos de Σ temos duas consequências principais: para $\Sigma \to 0$ temos,

pela forma da matriz de acoplamento, que os estados não mais estão acoplados, cada modo opera de forma individual; no caso em que $\Sigma \to \infty$ o efeito é de diminuir as flutuações do sistema e levar à um regime poissoniano. Para fazer mais clara a última afirmação, passamos para a próxima seção onde apresentamos as fórmulas para os momentos das distribuições de probabilidade em seu regime estacionário.

3.2 Valor médio, flutuação e fator de Fano

Os momentos das distribuições de probabilidade podem ser calculados através de derivadas das funções geratrizes, Eq.(2.5). Usando a relação de recorrência 3.12 e identificando os termos desta relação com a expansão em momentos do vetor função geratriz, Eq.3.6, temos que os momentos serão obtidos através de derivadas em relação à variável w avaliadas em $w = 0$. Assim sendo, para o vetor valor médio, temos a expressão

$$\left[\frac{d\vec{\phi}}{dw}\right]_{w=0} = \langle \vec{n} \rangle = \frac{\rho(\rho+\Sigma)1 + qP + \rho H}{\rho^2 + \rho\Sigma + q} N\vec{p}. \tag{3.13}$$

Cada componente do vetor valor médio corresponde à produção média de moléculas de mRNA de cada um dos estados operacionais do sistema. No entanto, o valor médio total é o mais significativo, uma vez que é esta quantidade que, comumente, é medida nos experimentos. O valor médio total é obtido somando se as componentes do vetor valor médio. De forma análoga, podemos obter esta quantidade considerando o produto escalar do autovetor à esquerda da matriz de acoplamento, $\vec{u} = (1, 1, 1)$, com o vetor $\langle \vec{n} \rangle$ ou seja,

$$\langle n \rangle = \vec{u}.\langle \vec{n} \rangle = \vec{u}\frac{\rho(\rho+\Sigma)1 + qP + \rho H}{\rho^2 + \rho\Sigma + q} N\vec{p}. \tag{3.14}$$

Considerando apenas o numerador da Eq.(3.14) temos que analisar três termos. O

primeiro envolve a matriz identidade e a matriz de eficiência, N, dando

$$\begin{aligned}\rho(\rho+\Sigma)\vec{u}N\vec{p} &= \rho(\rho+\Sigma)(1,1,1)\begin{pmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \\ &= \rho(\rho+\Sigma)(n_1p_1+n_2p_2+n_3p_3),\end{aligned} \tag{3.15}$$

o segundo termo, envolve a matriz P,

$$\begin{aligned}q\vec{u}PN\vec{p} &= q(1,1,1)\begin{pmatrix} p_1 & p_1 & p_1 \\ p_2 & p_2 & p_2 \\ p_3 & p_3 & p_3 \end{pmatrix}\begin{pmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \\ &= q(1,1,1)\begin{pmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \\ &= q(n_1p_1+n_2p_2+n_3p_3),\end{aligned} \tag{3.16}$$

lembrando que $p_1+p_2+p_3=1$. Por fim, o termo que envolve a matriz de acoplamento, H. Este é o termo mais simples de ser analisado uma vez que o vetor $\vec{u}$ é o autovetor à esquerda de H que corresponde ao autovalor nulo da mesma ($\vec{u}H=0$), e assim, temos:

$$\rho\vec{u}HN\vec{p}=0. \tag{3.17}$$

Combinando os resultados das Eqs.(3.15), (3.16) e (3.17) e substituindo na Eq.(3.14), temos para o valor médio total do sistema

$$\langle n\rangle = \frac{\rho(\rho+\Sigma)+q}{\rho(\rho+\Sigma)+q}(n_1p_1+n_2p_2+n_3p_3) = n_1p_1+n_2p_2+n_3p_3. \tag{3.18}$$

Analisando a fórmula para o valor médio total do sistema notamos que, ao contrário do que acontece com os valores médios parciais, Eq.(3.13), esta quantidade não depende do parâmetro Σ, ou seja, o valor médio total não será influenciado pelo fato do switch ser lento ou rápido. Como dissemos na seção anterior, temos dois regimes extremos para o parâmetro Σ. Para $\Sigma \to 0$ teremos o total desacoplamento do sistema e cada modo terá

valor médio dado pelas componentes do vetor

$$\lim_{\Sigma\to 0}\langle\vec{n}\rangle = \begin{pmatrix} n_1p_1 \\ n_2p_2 \\ n_3p_3 \end{pmatrix}. \tag{3.19}$$

No regime de switch rápido, $\Sigma \to \infty$, os valores médios parciais serão as componentes de

$$\lim_{\Sigma\to\infty}\langle\vec{n}\rangle = \langle n\rangle \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}. \tag{3.20}$$

Onde $\langle n\rangle$ é o valor médio total assim como dado pela Eq.(3.18).

Inspeção da expansão em momentos do vetor função geradora mostra que o vetor segundo momento pode ser calculado usando a relação de recorrência 3.12 pois

$$\frac{\langle\vec{n}^2\rangle - \langle\vec{n}\rangle}{2} = \vec{\phi}_2 = \frac{1}{2}\left[\frac{d^2\vec{\phi}}{dw^2}\right]_{w=0} \tag{3.21}$$

resolvendo para o vetor segundo momento teremos:

$$\left\langle\vec{n}^2\right\rangle = \left[\frac{d^2\vec{\phi}}{dw^2}\right]_{w=0} + \left[\frac{d\vec{\phi}}{dw}\right]_{w=0} = 2\vec{\phi}_2 + \vec{\phi}_1. \tag{3.22}$$

Nosso objetivo em se calcular o vetor segundo momento é utilizá-lo combinado com o vetor valor médio e analisar a flutuação do sistema. Para se obter a flutuação total do sistema, precisamos calcular o segundo momento total, $\langle n^2\rangle$, que é dado pelo produto escalar do vetor $\vec{u}$ com o vetor $\langle\vec{n}^2\rangle$,

$$\left\langle n^2\right\rangle = \vec{u}.\left\langle\vec{n}^2\right\rangle = 2\vec{u}.\vec{\phi}_2 + \vec{u}.\vec{\phi}_1. \tag{3.23}$$

O segundo termos da Eq.(3.23) já foi calculado e dá o valor médio total. O primeiro termos é:

$$2\vec{u}.\vec{\phi}_2 = \frac{4\rho^2 + 2\rho\Sigma + q}{4\rho^2 + 2\rho\Sigma + q}\vec{u}N\vec{\phi}_1 = \vec{u}N\vec{\phi}_1. \tag{3.24}$$

vamos olhar para o resultado acima com mais atenção

$$\vec{u}N\vec{\phi}_1 = \vec{u}N\frac{(\rho^2 + \rho\Sigma)1 + qP + \rho H}{\rho^2 + \rho\Sigma + q}N\vec{p}, \tag{3.25}$$

temos novamente três termos a serem analisados. O primeiro

$$\frac{(\rho^2+\rho\Sigma)\vec{u}N^2\vec{p}}{\rho^2+\rho\Sigma+q} = \frac{\rho^2+\rho\Sigma}{\rho^2+\rho\Sigma+q}(n_1^2p_1+n_2^2p_2+n_3^2p_3), \tag{3.26}$$

e o segundo termo fica:

$$\frac{q\vec{u}NPN\vec{p}}{\rho^2+\rho\Sigma+q} = \frac{q\left\langle n\right\rangle^2}{\rho^2+\rho\Sigma+q}. \tag{3.27}$$

O último termo da Eq.(3.25), que envolve a matriz H, não é tão simples de ser analisado como no caso anterior, quando olhamos para o valor médio total, devido ao fato da matriz N aparecer à esquerda de H. No entanto, para realizar este cálculo, vamos usar a matriz H em um formato mais conveniente explorando os parâmetros do espaço biológico do modelo. Temos

$$H = \frac{\Sigma}{\Delta}\begin{pmatrix} -1 & \epsilon_2 & (1-\epsilon_3) \\ (1-\epsilon_1) & -1 & \epsilon_3 \\ \epsilon_1 & (1-\epsilon_2) & -1 \end{pmatrix}\begin{pmatrix} \delta_1p_2p_3 & 0 & 0 \\ 0 & p_1\delta_2p_3 & 0 \\ 0 & 0 & p_1p_2\delta_3 \end{pmatrix}. \tag{3.28}$$

e podemos colocar a expressão de H da seguinte maneira:

$$H = \frac{\Sigma p_1p_2p_3}{\Delta}\epsilon\delta\pi^{-1} \tag{3.29}$$

onde a matriz δ é diagonal e tem a forma

$$\delta = \begin{pmatrix} \delta_1 & 0 & 0 \\ 0 & \delta_2 & 0 \\ 0 & 0 & \delta_3 \end{pmatrix}, \tag{3.30}$$

e a matriz

$$\pi = \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{pmatrix}. \tag{3.31}$$

Usando esta nova forma para a matriz H e o fato que

$$\begin{aligned} \vec{u}N &= (n_1,n_2,n_3) = \vec{n} \\ N\vec{p} &= \pi(n_1,n_2,n_3)^T = \pi\vec{n}^T, \end{aligned} \tag{3.32}$$

o último termo a ser calculado da Eq.(3.25) fica,

$$\frac{\rho\vec{u}NHN\vec{p}}{\rho^2+\rho\Sigma+q} = \frac{\rho\Sigma p_1p_2p_3}{\Delta(\rho^2+\rho\Sigma+q)}\vec{n}\epsilon\delta\vec{n}^T, \tag{3.33}$$

e o produto matricial é

$$\vec{n}\epsilon\delta\vec{n}^T = -[\delta_1 n_1^2 + \delta_2 n_2^2 + \delta_3 n_3^2 - n_1n_2\Gamma^3_{1,2} - n_1n_3\Gamma^2_{1,3} - n_2n_3\Gamma^1_{2,3}]. \tag{3.34}$$

Com as constantes Γ dadas como funções dos δ, e têm as expressões

$$\begin{aligned} \Gamma^3_{1,2} &= \delta_1+\delta_2-\delta_3, \\ \Gamma^2_{1,3} &= \delta_1+\delta_3-\delta_2, \\ \Gamma^1_{2,3} &= \delta_2+\delta_3-\delta_1. \end{aligned} \tag{3.35}$$

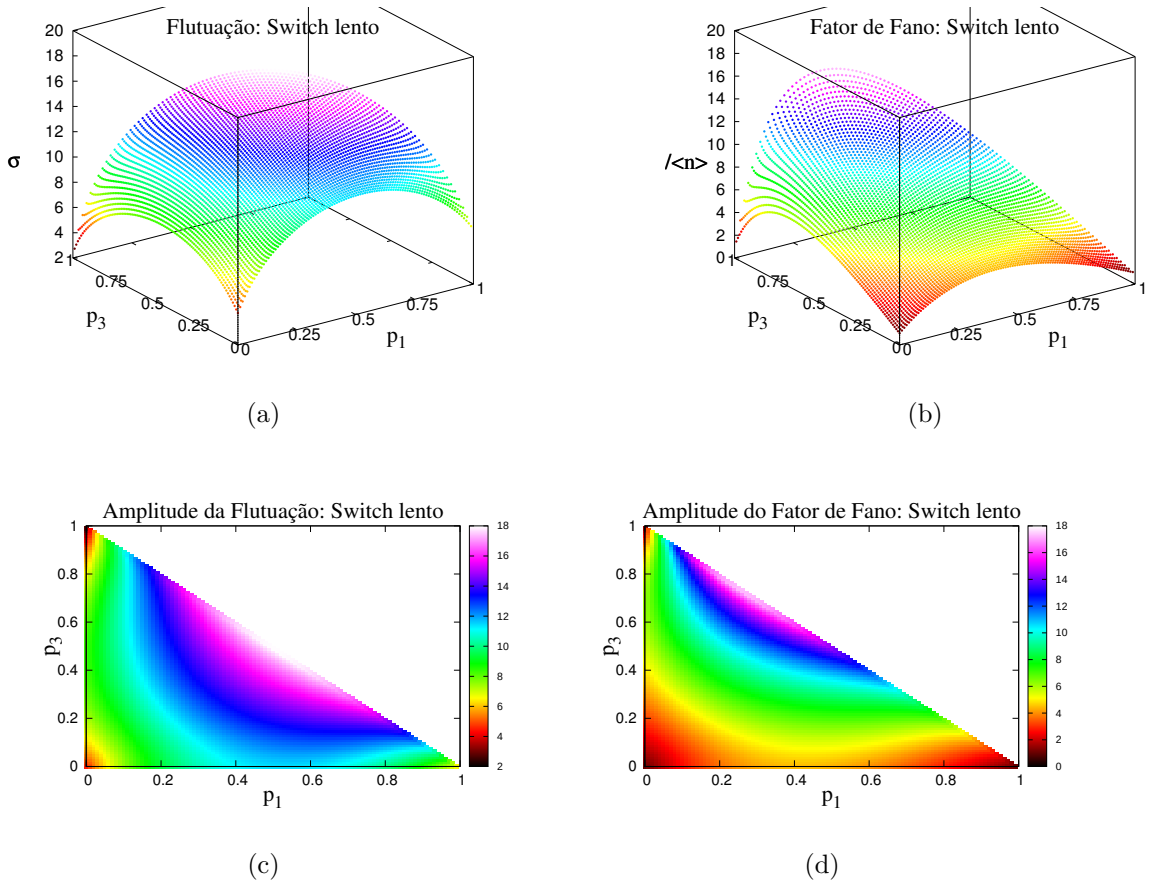

Figura 3.3 – Neste conjunto de gráficos nós exibimos o comportamento da flutuação total do sistema dado por: $\sigma = \sqrt{\langle n^2\rangle - \langle n\rangle^2}$, onde σ^2 é dado na Eq.(3.37) e também do fator de Fano, Eq.(3.38) no regime de switch lento, $\Sigma = 0.5$. As Figs.3.3(c) e 3.3(d) são os gráficos de densidade para as flutuações e para o fator de Fano, respectivamente. Para todos os gráficos temos: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = 0.5$ e $\rho = 1$.

Com os resultados das Eqs.(3.26), (3.27) e (3.33), temos o segundo momento total

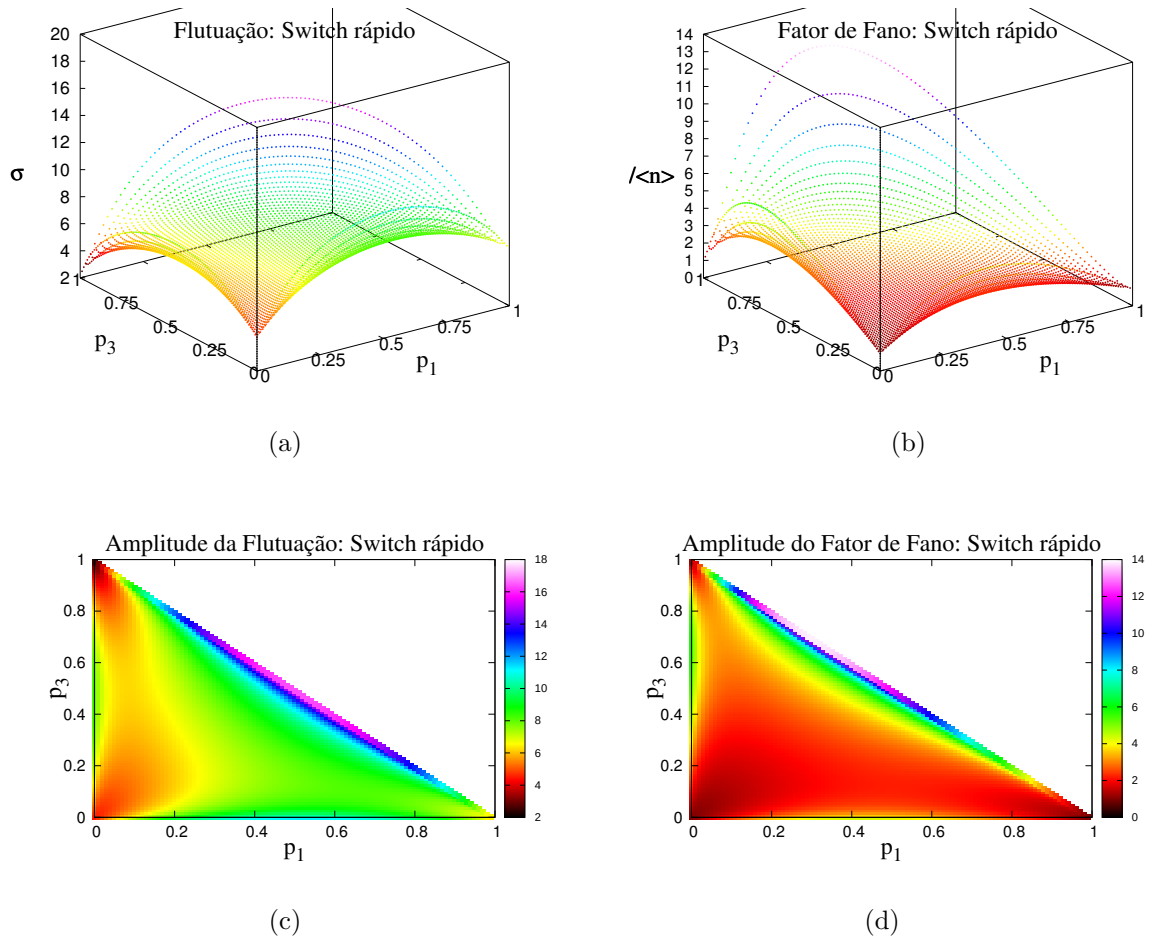

Figura 3.4 – Neste conjunto de gráficos nós exibimos o comportamento da flutuação total do sistema dado por: $\sigma = \sqrt{\langle n^2\rangle - \langle n\rangle^2}$, onde σ^2 é dado na Eq.(3.37) e também do fator de Fano, Eq.(3.38) no regime de switch rápido, $\Sigma = 5$. As Figs.3.4(c) e 3.4(d) são os gráficos de densidade para as flutuações e para o fator de Fano, respectivamente. Para todos os gráficos temos: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = 0.5$ e $\rho = 1$.

calculado e estamos em posição de olhar para a variância total do sistema que será dada pela expressão

$$\sigma^2 = \left\langle n^2 \right\rangle - \langle n \rangle^2 \tag{3.36}$$

que vai ter a forma explícita em termos dos parâmetros biológicos do modelo:

$$\begin{aligned} \sigma^2 &= \langle n \rangle + \frac{(\rho^2 + \rho\Sigma)[({n_1}^2 p_1 + {n_2}^2 p_2 + {n_3}^2 p_3) - (n_1 p_1 + n_2 p_2 + n_3 p_3)^2]}{\rho^2 + \rho\Sigma + q} + \\ &- \frac{\Sigma\rho p_1 p_2 p_3}{(\Delta\rho^2 + \rho\Sigma + q)}[\delta_1 n_1^2 + \delta_2 n_2^2 + \delta_3 n_3^2 - n_1 n_2 \Gamma^3_{1,2} - n_1 n_3 \Gamma^2_{1,3} - n_2 n_3 \Gamma^1_{2,3}]. \end{aligned} \tag{3.37}$$

A flutuação total do sistema será medida calculando se a raíz quadrada da variância do sistema.

Antes de discutir o comportamento da flutuação total do sistema apresentamos uma segunda quantidade comunmente utilizada para classificar o ruído e uma rede gênica.

Esta quantidade é dada pelo fator de Fano que tem a expressão:

$$\frac{\sigma^2}{\langle n \rangle} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle}. \tag{3.38}$$

O fator de Fano é utilizado para se medir o quão diferente uma distribuição de probabilidade é da distribuição de Poisson. Para valores do fato de Fano menores que a unidade dizemos que o sistema opera em um regime sub-poissoniano, maior que um é o regime super-poissoniano e quando o fator de Fano é idêntico à unidade a distribuição é poissoniana.

Com as fórmulas para a flutuação total do sistema e do fator de Fano em mãos podemos fazer alguns comentários sobre as características destes objetos. Quando o sistema opera em um único estado ($p_1 = 1$ ou $p_2 = 1$ ou $p_3 = 1$) pela Eq.(3.37) podemos ver que os dois últimos termos são idênticos a zero e a flutuação é igual ao valor médio. Com a ajuda da Eq.(3.38), vemos que, neste regime de um único modo operacional ativo, o fator de Fano é igual à um e a distribuição é poissoniana. O menor nível de ruído no sistema é atingido quando o este opera no estado de menor eficiência ($p_3 = 1$) com flutuação dada por $\sqrt{n_3}$. Sempre que uma das probabilidades de ocupação assintóticas for igual a zero o terceiro termo da Eq.(3.37) se anula e a flutuação é governada pelos termos quadráticos que aparecem no segundo termo da mesma equação e tem um formato parabólico. Notamos, também, que no denominador da Eq.(3.37) aparece um termo linear em q. Como mostramos na seção 2 do capítulo 2 desta tese, Eq.(2.30), a dependência deste parâmetro com Σ é quadrática. Como no numerador da Eq.(3.37) os termos em Σ são lineares, vemos que o ruído do sistema decai com $1/\Sigma$ quando $\Sigma \to \infty$ até o valor limite em que $\sigma^2 = \langle n \rangle$ ou, usando a Eq.(3.38), $\sigma^2 / \langle n \rangle = 1$, ou seja, quanto mais hábil for switch para atingir a sua configuração de equilíbrio mais próximo de uma distribuição de Poisson será a distribuição de probabilidade que descreve o sistema.

Como exemplo do comportamento da flutuação total do sistema e do fator de Fano, apresentamos dois conjuntos de gráficos, que foram obtidos usando se as Eqs.(3.37) e (3.38), para o regime de switch lento Fig.(3.3) e rápido Fig.(3.4). Os gráficos foram calculados usando como variáveis as probabilidades de ocupação dos estados, como indicado

nos eixos das figuras. Os elementos da matriz de eficiência são os mesmos da seção anterior ($n_1 = 40$, $n_2 = 20$ e $n_3 = 5$) e o esquema de acoplamento também foi mantido, sendo aquele onde todos os estados estão conectados e com mesma intensidade, dado pelos valores: $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1/2$.

Na Fig.3.3(a) mostramos a flutuação do sistema, na Fig.3.3(b) o fator de Fano e as duas restantes, neste conjunto de gráficos para switch lento, são os gráficos de densidade das superfícies que seguem os esquemas de cores à direita de cada um dos gráficos para mostrar a amplitude das funções. A região de maior flutuação é aquele em switch opera predominantemente nos seus estados extremos (1 e 3). Nota se que a possibilidade de ocupação do estado intermediário leva a redução das flutuações.

No caso da Fig.(3.4), em que o switch é rápido, nota se que a flutuação é reduzida em geral e que a distribuição vai ficando cada vez mais próxima de uma distribuição de Poisson como indica a Fig.3.4(b) e mais claro no gráfico de densidade para esta função Fig.3.4(d).

Como dito na discussão da Eq.(3.37) o esquema de menor flutuação do sistema é obtido quando este opera em seu estado de mais baixa eficiência. No entanto, o baixo nível de ruído neste modo operacional é acompanhado por um valor médio de moléculas produzidas, $\langle n \rangle = n_3$, da ordem de poucos unidades. O parâmetro n_3 representa a menor eficiência transcricional do gene quando comparado com os demais podendo, inclusive, ser zero no caso em que o estado basal não produz mRNA. A diminuição do ruído em sistemas biológicos tem que ser acompanhada de um número médio de partículas produzidas que sejam suficientes para formar complexos em concentrações mínimas para exercer sua função reguladora. Este fato é conhecido na literatura como o fenômeno de tudo ou nada (48) onde experimentos mostram que há uma concentração mínima, que uma classe de enzimas precisa atingir, de forma a se tornar ativa e realizar sua atividade no meio intracelular. Desta forma, vemos que para um sistema multimodal, como o aqui proposto com três estados de eficiência transcricional, uma boa estratégia de redução do ruído mantendo-se um nível de expressão relativamente alto é a possibilidade do sistema operar em seu estado intermediário com probabilidade de ocupação não desprezível. Este

efeito de redução do ruído pode ser visto, em especial, nas Figs.3.3(c) e 3.4(c) e o fato de o valor médio de partículas produzidas ser relativamente alto em um esquema de menor ruído pode ser observado nas Figs.3.3(d) e 3.4(d).

3.3 Distribuição de probabilidade com valor médio fixo

Nesta seção iremos analisar como a forma das distribuições de probabilidade e a amplitude do ruído do sistema se alteram quando este é obrigado a manter seu nível de produção médio fixo. Para tanto, usamos a Eq.(3.18) para fixar o valor médio de partículas no sistema e olhamos para os possíveis valores das probabilidades de ocupação e para ruído proveniente destas possíveis configurações do sistema. Com efeito, temos

$$\langle n \rangle_{fixo} = N = n_1 p_1 + n_2 p_2 + n_3 p_3, \tag{3.39}$$

onde $\langle n \rangle_{fixo} = N$ indica que o valor médio está fixado e seu valor é mantido igual à N. Primeiro, como $p_1 + p_2 + p_3 = 1$ podemos usar a Eq.(3.39) e resolver para p_1 e para p_3 obtendo duas expressões para estas quantidades como funções da probabilidade de ocupação do estado intermediário, p_2. Assim temos,

$$\begin{aligned} p_1 &= \frac{N - n_3}{n_1 - n_3} - p_2 \frac{n_2 - n_3}{n_1 - n_3}, \\ p_3 &= \frac{n_1 - N}{n_1 - n_3} - p_2 \frac{n_1 - n_2}{n_1 - n_3}. \end{aligned} \tag{3.40}$$

A necessidade de se obter expressões tanto para p_1 quanto para p_3 em termos de p_2 vem do fato destes parâmetros serem limitados com domínio entre zero e um. Dependendo do valor escolhido para a produção média do sistema, N, teremos que analisar qual o possível domínio de variação de p_2 de forma a não ultrapassarmos os valores extremos. Fazendo $p_1 = 0$ na primeira das equações em (3.40) e resolvendo para p_2 temos que p_2 será uma função de N e o mesmo pode ser concluído fazendo $p_3 = 0$ na segunda equação. Obtemos

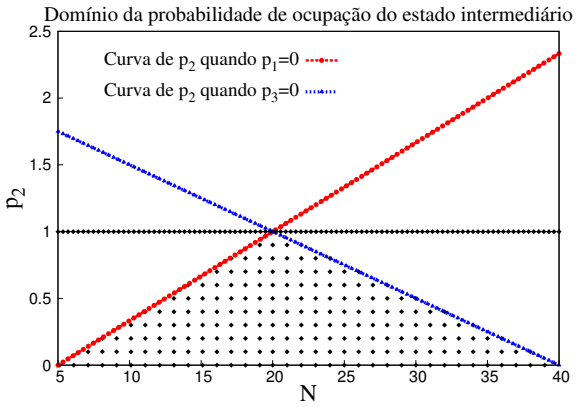

Figura 3.5 – Nesta figura mostramos o domínio da probabilidade de ocupação do estado intermediário como função do valor médio fixo N. Os parâmetros de eficiência transcricional foram escolhidos como anteriormente e têm o valor: $n_1 = 40$, $n_2 = 20$ e $n_3 = 5$.

assim, as expressões,

$$p_2 = \frac{N - n_3}{n_2 - n_3} \text{ ou } p_2 = \frac{n_1 - N}{n_1 - n_2} \tag{3.41}$$

Analisando estas duas expressões para p_2 como função de N notamos que o máximo valor que este parâmetro poderá assumir, uma vez que o valor médio está fixo, será o mínimo valor entre as duas funções. Esta discussão está exemplificada na Fig.(3.5) onde mostramos o gráfico de p_2 como função de N, usando as duas equações de 3.41, como indicado na figura. Seguindo as seções anteriores os valores para os parâmetros de eficiência transcricional foram escolhidos como: $n_1 = 40$, $n_2 = 20$ e $n_3 = 5$. A região preenchida pelos pontos e encerrada pelas curvas de p_2 quando $p_1 = 0$ (vermelha) e $p_3 = 0$ (azul) é região de valores possíveis para p_2. Notamos que quando $N = n_3 = 5$ a única configuração possível para o sistema é de $p_3 = 1$ e quando $N = n_1 = 40$ a única configuração possível é de $p_1 = 1$. No caso em que o valor médio é mantido igual a eficiência do estado intermediário, $N = n_2 = 20$, p_2 pode varia de zero à um e todas as configurações do sistema são atingíveis. Usamos este último valor de N para analisar as diferentes distribuições de probabilidade que descrevem o sistema e o ruído associado.

Na Fig.3.6(a) temos um gráfico de densidade de cor para a amplitude da flutuação e notamos que para qualquer valor médio fixo a existência da possibilidade do sistema operar no estado intermediário ($p_2 \neq 0$) leva a atenuação do ruído total. No caso em que p_2 pode variar de zero à um (para $< n >= 20$) observa se que a possibilidade de popular o estado intermediário leva as flutuações de um valor próximo de 20 ($p_2 = 0$)

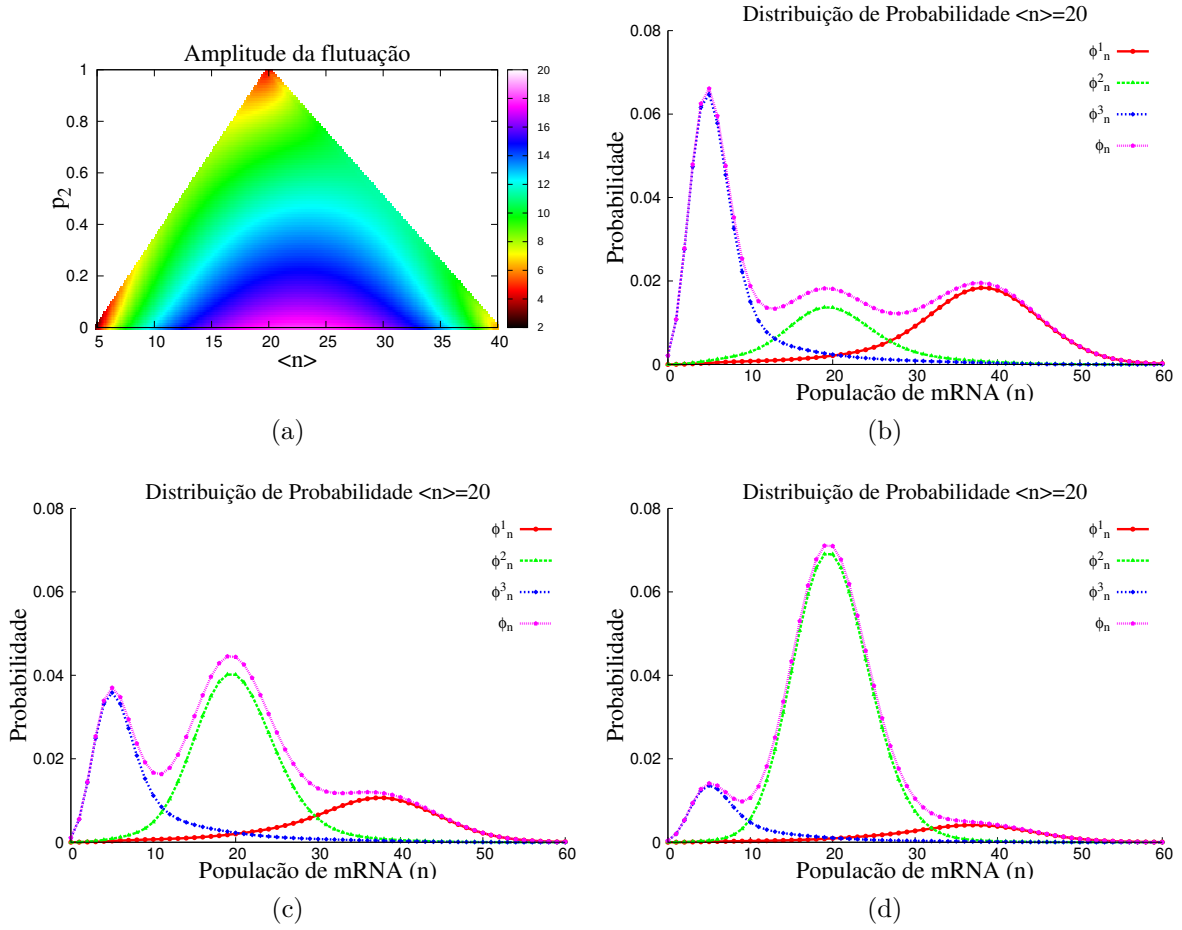

Figura 3.6 – Neste conjunto de gráficos mostramos o gráfico de densidade para a amplitude da flutuação como função de $< n >$ e p_2, Fig.3.6(a). O restante dos gráficos mostra o comportamento das distribuições de probabilidade parciais e total no regime de switch lento, $\Sigma = 0.5$, valor médio fixo, $< n >= 20$, e diferentes valores das probabilidades de ocupação. Na Fig.3.6(b) $p_1 = 0.35, p_2 = 0.2$ e $p_3 = 0.45$, na Fig.3.6(c) $p_1 = 0.22$, $p_2 = 0.5$ e $p_3 = 0.28$ e na Fig.3.6(d) $p_1 = 0.09, p_2 = 0.8$ e $p_3 = 0.11$. Para todos os gráficos: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = 0.5$ e $\rho = 1$.

para um valor próximo de 4 quando $p_2 = 1$. Os demais gráficos mostram as diferentes formas das distribuições de probabilidade mas com mesmo valor médio. Na Fig.3.6(b) temos a situação de maior ruído que transiciona para esquemas de menor flutuação com o aumento na probabilidade de popular o estado intermediário Fig.3.6(c) e Fig.3.6(d), respectivamente.

Capítulo 4

Dinâmica dos momentos

Neste capítulo, iremos apresentar as equações dependentes do tempo para os momentos das distribuições de probabilidade. Iremos, em um primeiro movimento, resolver as equações para os três primeiros momentos: de ordem zero, primeira e segunda ordem que irão descrever, respectivamente, as dinâmicas para as probabilidades de ocupação, valores médios parciais e flutuações. Em seguida, apresentaremos o comportamento dinâmico destas quantidades para diferentes acoplamentos e discutimos as estratégias que o sistema pode adotar para alcançar sua configuração assintótica.

4.1 A solução dependente do tempo para os momentos das distribuições

Os momentos das distribuições de probabilidade podem ser obtidos através de derivadas das funções geradoras assim como posto na Eq.(2.5) ou, equivalentemente, na Eq.(2.12). Escolhemos usar as equações em sua forma vetorial. Para o vetor momento de ordem zero, cujo as componentes são as probabilidades de ocupação de cada estado, as equações

são aquelas apresentadas no Capítulo 2 desta tese, Eq.(2.15), e são:

$$\frac{d\vec{P}(t)}{dt} = H\vec{P}(t). \tag{4.1}$$

Usando a fórmula 2.12 com $j = 1$ e aplicando na Eq(2.10) obtemos para o vetor valor médio a equação diferencial ordinária,

$$\frac{d\left\langle \vec{n} \right\rangle (t)}{dt} = (H - \rho)\left\langle \vec{n} \right\rangle (t) + K\vec{P}(t), \tag{4.2}$$

onde as matrizes K e ρ são as mesmas apresentadas na primeira seção do Capítulo 2 desta tese. Colocando $j = 2$ em 2.12 e atuando novamente na Eq.(2.10) obtemos para o vetor segundo momento

$$\frac{d\left\langle \vec{n}^2 \right\rangle (t)}{dt} = (H - 2\rho)\left\langle \vec{n}^2 \right\rangle (t) + 2K\left\langle \vec{n} \right\rangle (t). \tag{4.3}$$

Nota se que a estrutura das equações diferenciais ordinárias para o momento de ordem j apresenta a matriz H subtraída de j vezes a matriz ρ e vezes o próprio momento considerado e um termo heterogênio que envolve a matriz K vezes j vezes o momento de ordem $j - 1$. Desta forma, a equação diferencial para um momento de ordem j qualquer fica:

$$\frac{d\left\langle \vec{n}^j \right\rangle (t)}{dt} = (H - j\rho)\left\langle \vec{n}^j \right\rangle (t) + jK\left\langle \vec{n}^{j-1} \right\rangle (t). \tag{4.4}$$

A estrutura da equação diferencial ordinária para o j-ésimo momento da distribuição e de uma EDO linear de primeira ordem com um termo heterogêneo. Para se calcular sua solução geral propomos o ansatz

$$\left\langle \vec{n}^j \right\rangle (t) = e^{(H-j\rho)t}\vec{\eta}(t) \tag{4.5}$$

substituindo na Eq.(4.4) chegamos a seguinte EDO para $\vec{\eta}(t)$:

$$\frac{d\vec{\eta}(t)}{dt} = e^{-(H-j\rho)t} jK\left\langle \vec{n}^{j-1} \right\rangle (t). \tag{4.6}$$

Por integração direta da Eq.(4.6) obtemos,

$$\vec{\eta}(t) = \vec{\eta}(0) + j\int_0^t e^{-(H-j\rho)\bar{t}} jK\left\langle \vec{n}^{j-1} \right\rangle (\bar{t})d\bar{t}. \tag{4.7}$$

Usando a Eq.(4.5) a solução geral para o j-ésimo momento dependente do tempo fica:

$$\left\langle \vec{n}^j \right\rangle (t) = e^{(H-j\rho)t} \left\langle \vec{n}^j \right\rangle (0) + j \int_0^t e^{(H-j\rho)(t-\bar{t})} K \left\langle \vec{n}^{j-1} \right\rangle (\bar{t}) d\bar{t}. \tag{4.8}$$

As exponenciais que aparecem na solução de $\langle \vec{n}^j \rangle (t)$ são exponenciais de matrizes e em geral não são simples de se calcular. No caso em que estamos tratando temos as matrizes H e $j\rho$ como argumento das exponenciais. A matriz $j\rho$ é múltipla da matriz identidade e, assim, comuta com a matriz H,

$$[H, j\rho] = Hj\rho - j\rho H = 0, \tag{4.9}$$

e devido ao fato destas matrizes comutarem podemos escrever a exponencial da soma das matrizes $H - j\rho$ como o produto de duas exponenciais,

$$e^{(H-j\rho)t} = e^{Ht} e^{-j\rho t}. \tag{4.10}$$

Como dito acima, a matriz $j\rho$ é múltipla da identidade e, portanto, diagonal. A exponencial de uma matriz diagonal é fácil de se calcular, basta exponencializar os elementos da diagonal. Já a matriz H não é diagonal e para fazer a operação de exponencialização usaremos uma transformação de similaridade onde

$$H = SDS^{-1}. \tag{4.11}$$

A matriz S é construída a partir dos autovetores da matriz H, que foram calculados na seção 2 do capítulo 2, e tem a forma

$$S = \begin{pmatrix} p_1 & 1 & 1 \\ p_2 & A^+ & A^- \\ p_3 & -1 - A^+ & -1 - A^- \end{pmatrix}, \tag{4.12}$$

sendo $A^\pm$ como definidos na Eq.(2.35). A matriz S^{-1} é a inversa da matriz S e tem a forma

$$S^{-1} = \begin{pmatrix} A^+ - A^- & A^+ - A^- & A^+ - A^- \\ -A^-(p_2 + p_3) - p_2 & (1 + A^-)p_1 + p_3 & A^- p_1 - p_2 \\ A^+(p_2 + p_3) + p_2 & -(1 + A^+)p_1 - p_3 & -A^+ p_1 + p_2 \end{pmatrix}. \tag{4.13}$$

A matriz D é diagonal e seus elementos são os autovalores da matriz H, também calculados na seção 2 do capítulo 2. A matriz D é:

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda^+ & 0 \\ 0 & 0 & \lambda^- \end{pmatrix}. \tag{4.14}$$

Usando a transformação de similaridade, obtemos

$$e^{Ht} = e^{(SDS^{-1})t}. \tag{4.15}$$

Expandindo a exponencial da direita em série de Taylor da variável t em torno de $t = 0$ temos

$$e^{(SDS^{-1})t} = 1 + SDS^{-1}t + SDS^{-1}SDS^{-1}\frac{t^2}{2!} + \mathrm{O}(t^3). \tag{4.16}$$

Observando que $SS^{-1} = 1$ podemos escrever a expansão acima, da forma

$$e^{(SDS^{-1})t} = S[1 + Dt + D^2\frac{t^2}{2!} + \mathrm{O}(t^3)]S^{-1} = Se^{Dt}S^{-1}. \tag{4.17}$$

Com este resultado em mãos, a solução geral para os momentos das distribuições é dada como

$$\left\langle \vec{n}^j \right\rangle (t) = Se^{Dt}S^{-1}e^{-j\rho t}\left\langle \vec{n}^j \right\rangle (0) + j\int_0^t Se^{D(t-\bar{t})}S^{-1}e^{-j\rho(t-\bar{t})}K\left\langle \vec{n}^{j-1} \right\rangle (\bar{t})d\bar{t}. \tag{4.18}$$

Mais uma vez, como a matriz $j\rho$ é múltipla da identidade irá comutar com as matrizes S e D, a expressão final para a solução geral dos momentos fica:

$$\left\langle \vec{n}^j \right\rangle (t) = Se^{(D-j\rho)t}S^{-1}\left\langle \vec{n}^j \right\rangle (0) + j\int_0^t Se^{(D-j\rho)(t-\bar{t})}S^{-1}K\left\langle \vec{n}^{j-1} \right\rangle (\bar{t})d\bar{t}. \tag{4.19}$$

A matriz $e^{(D-j\rho)t}$ é diagonal e tem a forma:

$$D - j\rho = \begin{pmatrix} e^{-j\rho t} & 0 & 0 \\ 0 & e^{(\lambda^+ - j\rho)t} & 0 \\ 0 & 0 & e^{(\lambda^+ - j\rho)t} \end{pmatrix}. \tag{4.20}$$

Em particular, iremos olhar para os momentos com $j = 0, 1$ e 2 pois o nosso interesse é de analisar a variação dinâmica as probabilidade de ocupação (j=0), do valor médio

(j=1), da flutuação (que vai depender de j=2) e do fator de Fano. Antes de passar para a próxima seção, onde faremos uma discussão sobre os autovalores da matriz H, um pequeno comentário sobre as integrais que aparecem na solução geral. Para $j = 0$, temos a solução dinâmica das probabilidade de ocupação e notamos que elas não dependem de ρ e a integral não contribui, de forma que temos

$$\vec{P}(t) = Se^{Dt}S^{-1}\vec{P}(0). \tag{4.21}$$

A solução na Eq.(4.21) será o termo heterogêneo (que aparece dentro da integral) da solução para $j = 1$. Assim sendo, as integrais que teremos que efetuar são de funções exponenciais para todos os momentos e este tipo de integral é bastante simples de ser calculada.

4.2 Autovalores: regiões oscilantes

Nesta seção iremos fazer uma discussão sobre a forma dos autovalores da matriz H e analisar seu comportamento para alguns esquemas de acoplamento entre os estados. Para tanto, vamos reescrever as expressões para os autovalores não nulos da matriz H,

$$\lambda^{\pm} = -\frac{\Sigma}{2}\left(1 \pm \sqrt{1 - \frac{4\delta_1\delta_2\delta_3 p_1 p_2 p_3}{\Delta^2}}\right). \tag{4.22}$$

Os δ_i são os cofatores da matriz ϵ, Eq.(2.22), e dependem dos acoplamentos entre os estados e Δ é dado na Eq.(2.28). Vamos analisar o termo

$$\omega = 1 - \frac{4\delta_1\delta_2\delta_3 p_1 p_2 p_3}{\Delta^2}. \tag{4.23}$$

Notamos que sempre que qualquer um dos $\delta_i = 0$ implicará em $\omega = 1$ e o mesmo acontece quando qualquer uma das probabilidades de ocupação se anula. No entanto, existem acoplamentos e configurações do sistema que fazem com que $\omega < 0$ e então teremos estados oscilantes. Para mostrar este fato, exibimos na Fig.(4.1) quadro gráficos de densidade onde plotamos a amplitude de ω como função das probabilidades de ocupação

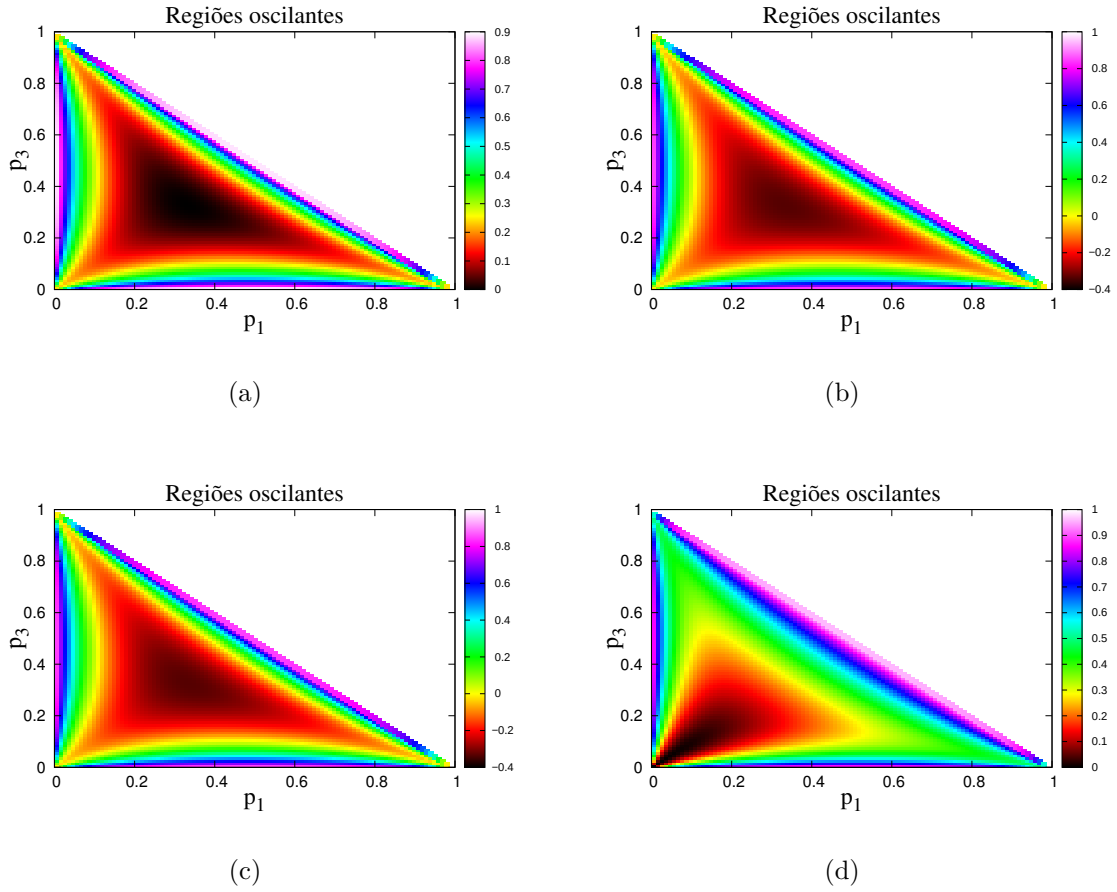

Figura 4.1 – Neste conjunto de gráficos apresentamos o gráfico de densidade para a quantidade $\omega = 1 - 4q/\Sigma^2$, como função de p_1 e p_3 e para diferentes esquemas de acoplamento. Na Fig.2.13, temos $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1/2$, na Fig.4.1(b) temos $\epsilon_1 = 1, \epsilon_2 = 1$ e $\epsilon_3 = 1$, na Fig.4.1(c) temos $\epsilon_1 = 0, \epsilon_2 = 0$ e $\epsilon_3 = 0$ e na Fig.4.1(d) temos $\epsilon_1 = 0, \epsilon_2 = 1/2$ e $\epsilon_3 = 1$.

e para quatro esquemas de acoplamento distintos.

Na Fig.4.1(a) os valores são $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1/2$, ou seja, todos os estados estão conectados e com mesma intensidade de acoplamento. Podemos ver que neste esquema a dinâmica do sistema é descrita por um decaimento exponencial uma vez que o menor valor que ω assume é zero. Na Fig.4.1(b) escolhemos o caso em que o acoplamento é anticíclico, $\epsilon_1 = 1, \epsilon_2 = 1$ e $\epsilon_3 = 1$, e para uma grande região do domínio de p_1 e p_3 (encerrada pela borda amarela) temos oscilações, ou seja, $\omega < 0$ e o mesmo se observa na Fig.4.1(c) onde o esquema é de acoplamento cíclico com $\epsilon_1 = 0, \epsilon_2 = 0$ e $\epsilon_3 = 0$. Na Fig.4.1(d) os elementos da matriz ϵ assumem os valores: $\epsilon_1 = 0, \epsilon_2 = 1/2$ e $\epsilon_3 = 1$. Neste caso, o estado 1 está conectado com o estado 2 mais não com o estado 3; e o 3 com o 2 mas não com o 1; e o 2 está conectado com o 1 e com o 3; e neste esquema de acoplamento não temos uma região de oscilações. Outros esquemas de acoplamento, que serão usados na próxima seção, são

quando o sistema opera com apenas dois de seus estados acoplados e o terceiro não tem comunicação com estes. Neste caso, também não temos oscilações pois sempre um dos δ_i será igual à zero.

Olhando para os casos em que há regiões de estados oscilantes, notamos que mesmo com $\omega < 0$ a sua amplitude é muito pequena, chegando ao máximo valor de -0.4, o que implica em uma baixa frequência de oscilação. Antes que o sistema oscile ele já decaiu e assim, como veremos na seção seguinte, não é observado durante a dinâmica dos momentos um comportamento oscilatório.

4.3 Comportamento dinâmico dos momentos para distintos esquemas de acoplamento

Nesta seção apresentamos o comportamento dinâmico das probabilidades de ocupação, dos valores médios parciais, das flutuações de cada estado e também a dinâmica do fator de Fano. Para cada conjunto de figuras escolhemos valores para os elementos da matriz ϵ distintos, indicados nas figuras, e assim explorando diversas estratégias de acoplamento que podem ser utilizadas pelo sistema. Em todas as figuras, os painéis da coluna da esquerda são para switch lento, $\Sigma = 0.5$, e os da direita para switch rápido, $\Sigma = 5$. Os elementos da matriz de eficiência, N, foram escolhido como anteriormente e têm os valores: $n_1 = 40$, $n_2 = 20$ e $n_3 = 5$.

As condições iniciais, para todas as figuras, são aquelas que correspondem ao vetor probabilidade de ocupação em $t = 0$,

$$\vec{P}(0) = (1/10, 8/10, 1/10)^T, \tag{4.24}$$

ou seja, em $t = 0$ o sistema opera predominantemente em seu estado de produção intermediária. Os demais valores de condições iniciais (valores médios e flutuações) são obtidos, substituindo o valor dos elementos da matriz acoplamento escolhidos, os parâ-

metros de eficiência e $\vec{P}(0)$ nas Eqs.(3.13) e (3.22). Como estado assintótico escolhemos aquele rotulado pelo vetor:

$$\vec{p} = (98/100, 1/100, 1/100)^T. \tag{4.25}$$

A escolha dos vetores $\vec{P}(0)$ e $\vec{p}$, tendo o óperon *lac* em mente, correspondem ao caso em que o sistema está em um meio com altas concentrações de glicose e lactose e com o passar do tempo a glicose é totalmente consumida sobrando apenas a lactose como fonte de energia. Assim, ele opera em seu estado intermediário, em $t = 0$, possibilitando o consumo tanto de glicose quanto de lactose e, no equilíbrio, quando não mais existe glicose, este está operando predominantemente em seu estado de maior eficiência pois há abundância de lactose no meio celular.

Em todas as figuras a primeira característica que pode ser observada é a diferença entre o caso de switch lento (painéis da esquerda) e rápido (painéis à direita) mostrando que Σ de fato é a componente principal que governa a dinâmica do sistema até atingir a configuração assintótica.

Nas Figs.(4.2), (4.3) e (4.4) os acoplamentos são: $\epsilon_1 = 1/2$, $\epsilon_2 = 1/2$ e $\epsilon_3 = 1/2$, $\epsilon_1 = 0.01$, $\epsilon_2 = 0.01$ e $\epsilon_3 = 0.01$e $\epsilon_1 = 0.99$, $\epsilon_2 = 0.99$ e $\epsilon_3 = 0.99$, respectivamente. No caso da Fig.(4.2) todos os estados estão conectados e com mesma intensidade, na Fig(4.3), o acoplamento é o cíclico, $1 \rightharpoonup 2 \rightharpoonup 3 \rightharpoonup 1$ e na Fig.(4.4) temos o acoplamento anticíclico, $1 \rightharpoonup 3 \rightharpoonup 2 \rightharpoonup 1$. Como, no instante $t = 0$ o sistema opera predominantemente em seu estado intermediário ($\vec{P}(0) = (1/10, 8/10, 1/10)^T$) e assintoticamente estará exclusivamente produzindo com máxima eficiência ($\vec{p} = (98/100, 1/100, 1/100)^T$) haverá um fluxo de probabilidade saindo do estado 2 e chegando no estado 1 para assim construir o estado de equilíbrio desejado. No caso em que o acoplamento é igual entre todos os estados notamos que o decréscimo da probabilidade de encontrarmos o sistema em seu estado intermediário aumenta de imediato a ocupação do estado 1 e discretamente a do estado 3. Quando o acoplamento é cíclico, ou seja, para transferir a probabilidade de ocupação do estado 2 para o estado 1 é necessário passar pelo estado 3, notamos que a ocupação do estado 3 aumenta, ficando maior que a do estado 1 até o ponto de igual probabilidade

para todos os estados e então decai o estado 2 e por transferência de probabilidade de ocupação do 3 para o 1 este, assintoticamente, acaba populado. No caso do acoplamento anticíclico a transferência de probabilidade se dá diretamente do estado 2 para o estado 1 e o estado 3 não tem contribuição significativa neste processo e sua probabilidade de ocupação decai monotonicamente.

Na Fig.(4.5) o acoplamento é dado por: $\epsilon_1 = 0.01$, $\epsilon_2 = 0.5$ e $\epsilon_3 = 0.99$. Ou seja, a transferência de probabilidade se dá do estado 1 para o estado 2 e aí para o estado 3. No entanto, entre o estado 2 e o estado 1 há fluxo de probabilidade e então o que observamos é uma pequena transferência do 2 para o 3 antes de total transferência do 2 para o 1.

Nas últimas três figuras que compõem a dinâmica dos momentos, escolhemos acoplamentos onde dois estados se comunicam com bastante intensidade e um deles está acoplado fracamente aos demais. Temos assim, nas Figs. (4.6), (4.7) e (4.8) os acoplamentos: $\epsilon_1 = 0.99$, $\epsilon_2 = 1/2$ e $\epsilon_3 = 0.01$, forte comunicação entre os estados 1 e 3; $\epsilon_1 = 0.5$, $\epsilon_2 = 0.01$ e $\epsilon_3 = 0.99$, inteso acoplamento entre os estados 2 e 3 e por fim, $\epsilon_1 = 0.01$, $\epsilon_2 = 0.99$ e $\epsilon_3 = 0.5$ estando fortemente acoplados os estados 1 e 2. Na Fig.(4.6) notamos que o sistema dificilmente atinge a configuração de equilíbrio uma vez que não há um acoplamento eficiente para transferir probabilidade de ocupação do estado intermediário para os demais. Na Fig.(4.7) o sistema terá que passa um tempo enorme ($t >> \rho$) para atingir a configuração assintótica uma vez que a transferência de informação se dá entre os estados 1 e 2. Na Fig.(4.8) o sistema tem forte comunicação entre os estados 1 e 2 transicionando com eficiência para a configuração de equilíbrio e mantendo o estado 3 se modificações.

No caso dos momentos de ordem superior toda a diferença na dinâmica está ligada às diferentes estratégias de acoplamento uma vez que as condições iniciais e assintóticas são as mesmas para todas as figuras. Selecionando um acoplamento o sistema pode ter maior ou menor nível de ruído e aproximar se ou distanciar se de estruturas poissonianas revelando uma flexibilidade grande e desejada para adaptação e evolução do sistema.

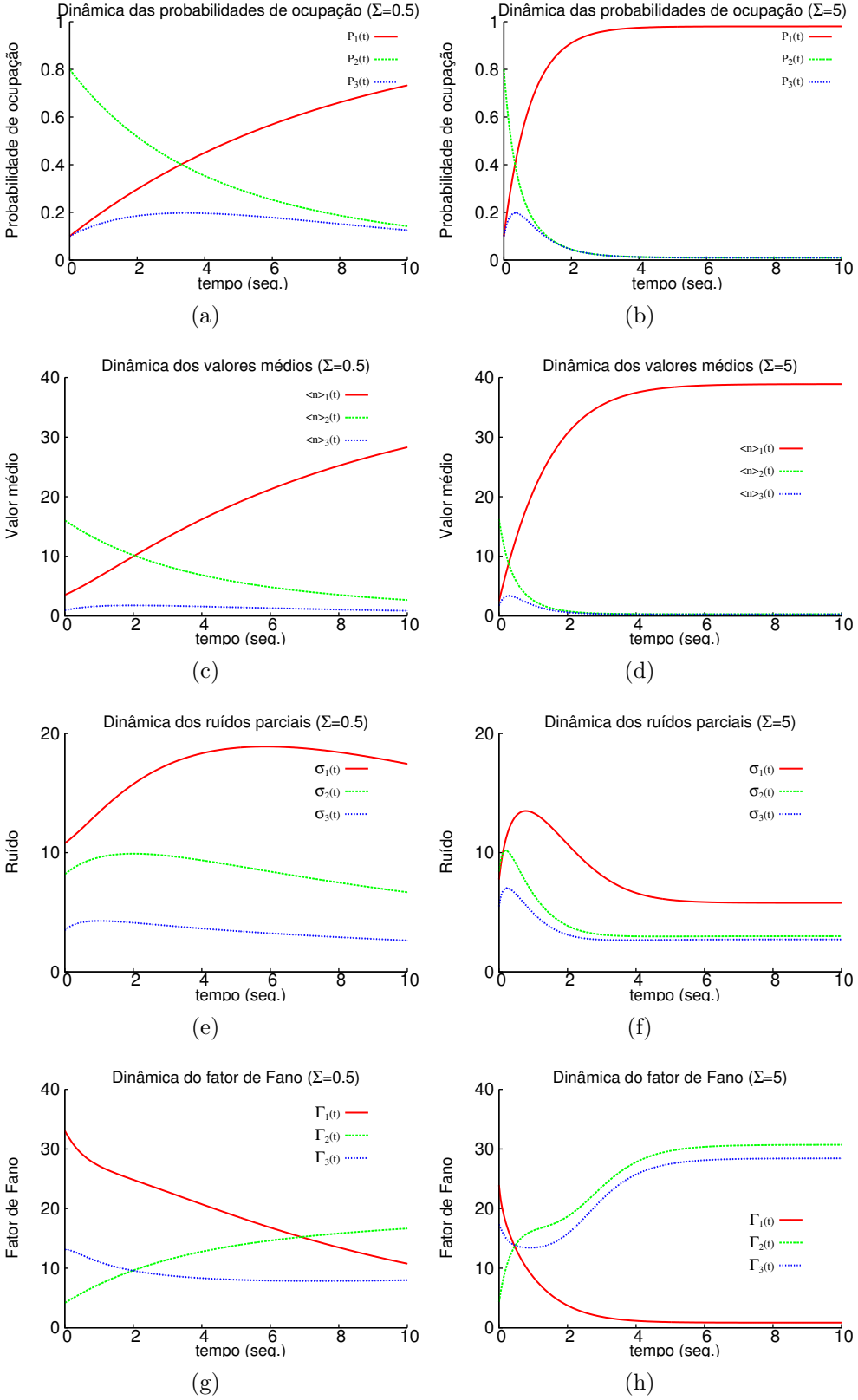

Figura 4.2 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 1/2$, $\epsilon_2 = 1/2$ e $\epsilon_3 = 1/2$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

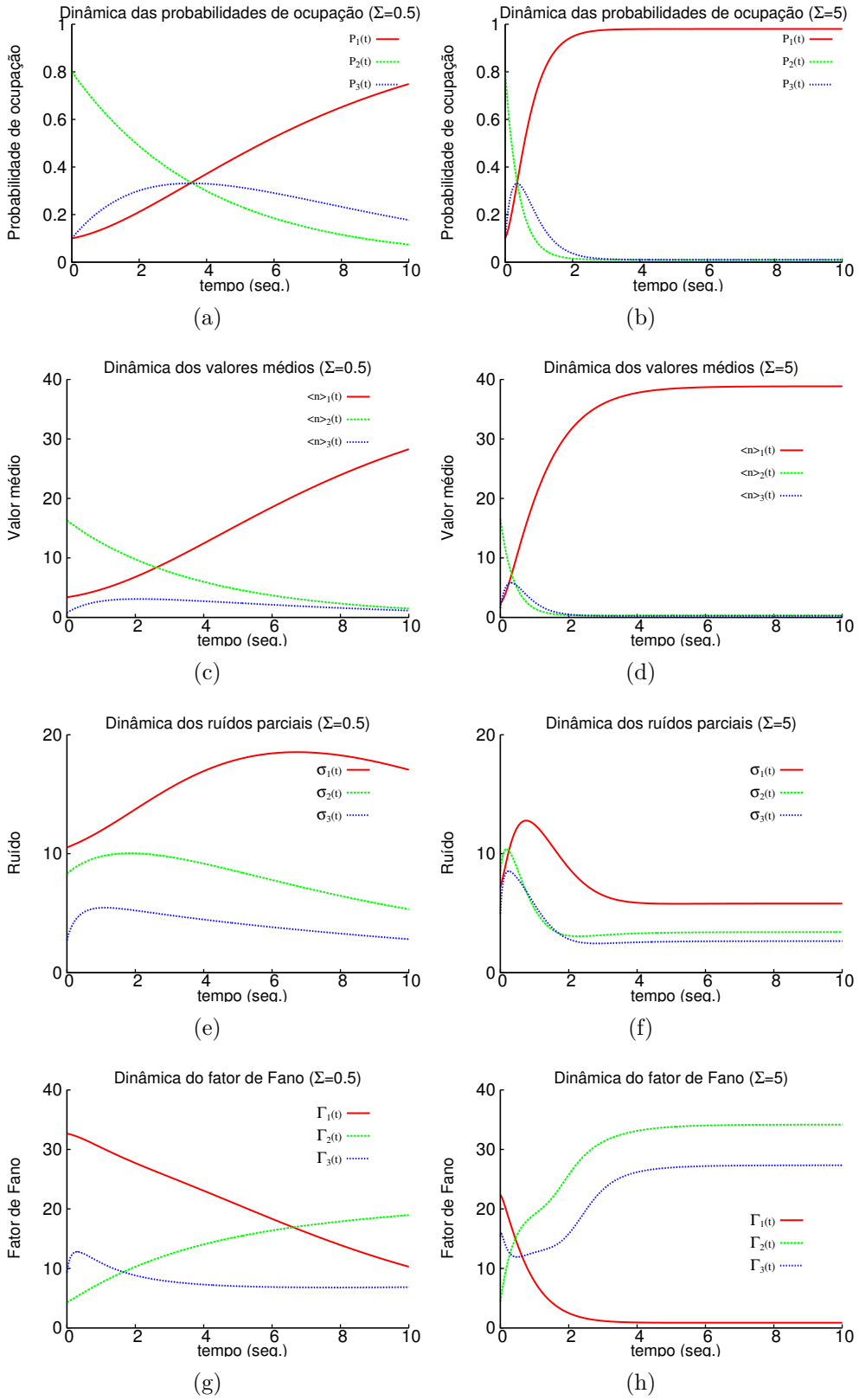

Figura 4.3 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 0.01$, $\epsilon_2 = 0.01$ e $\epsilon_3 = 0.01$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

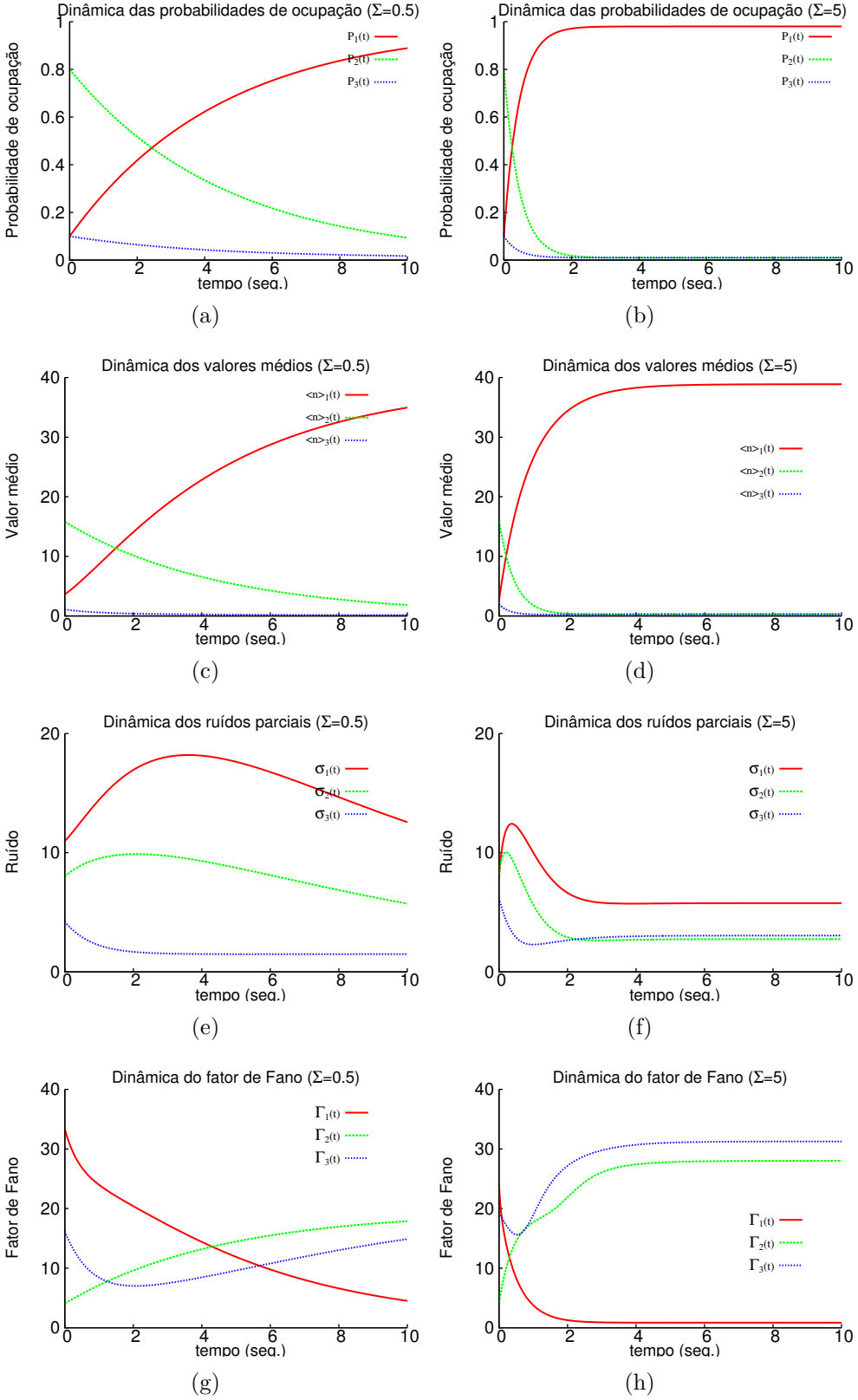

Figura 4.4 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 0.99$, $\epsilon_2 = 0.99$ e $\epsilon_3 = 0.99$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

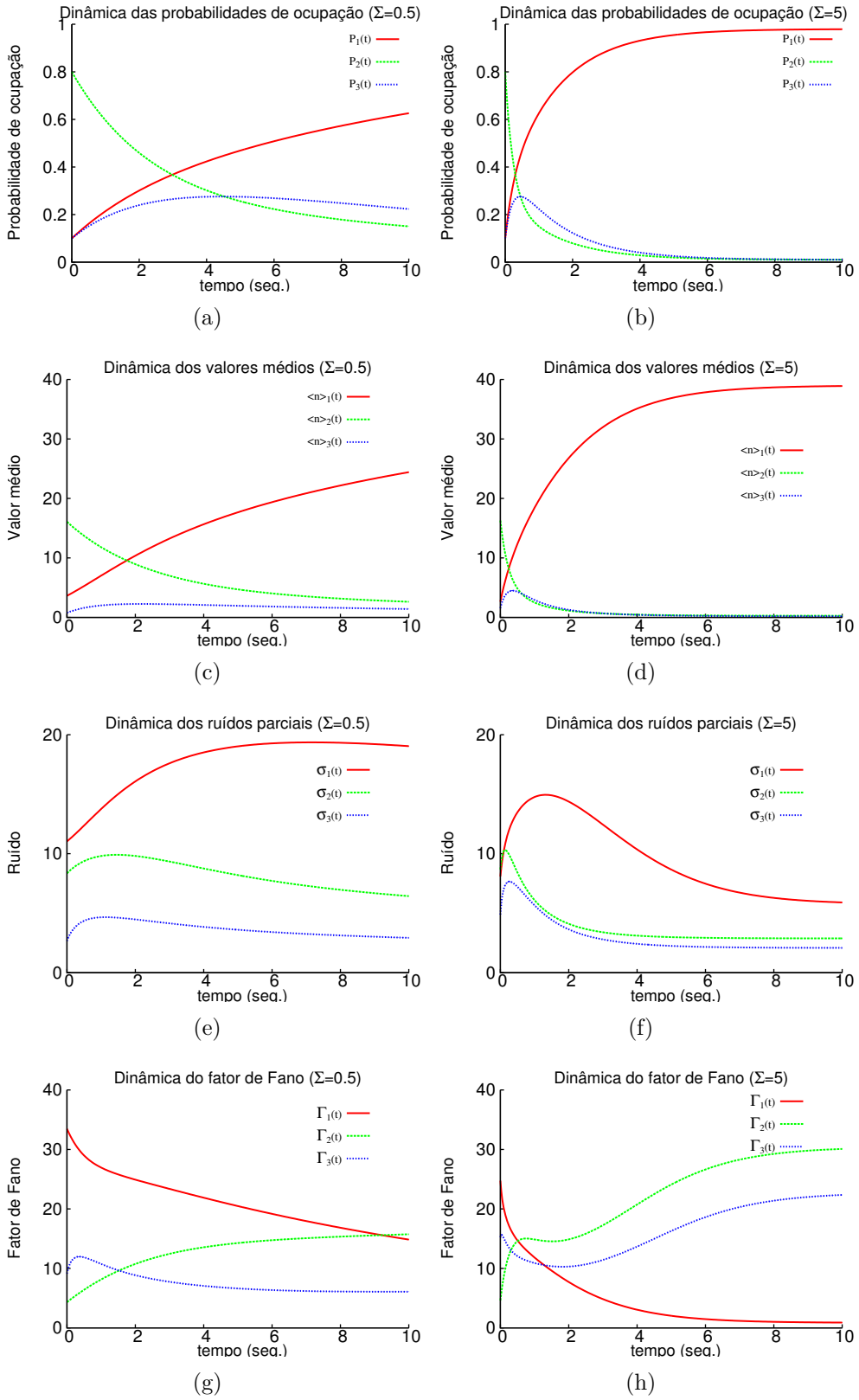

Figura 4.5 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 0.01$, $\epsilon_2 = 1/2$ e $\epsilon_3 = 0.99$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

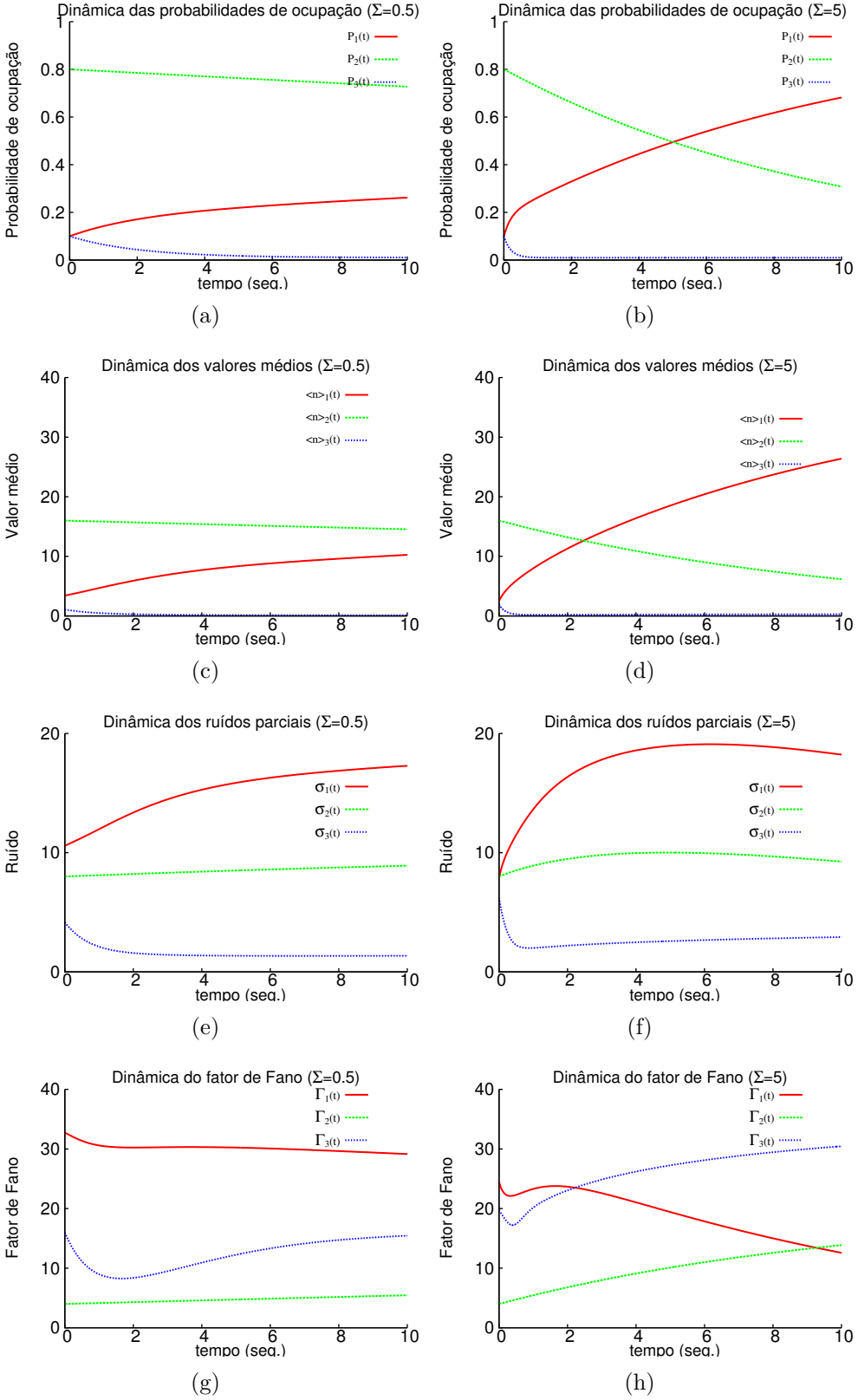

Figura 4.6 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 0.99$, $\epsilon_2 = 1/2$ e $\epsilon_3 = 0.01$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

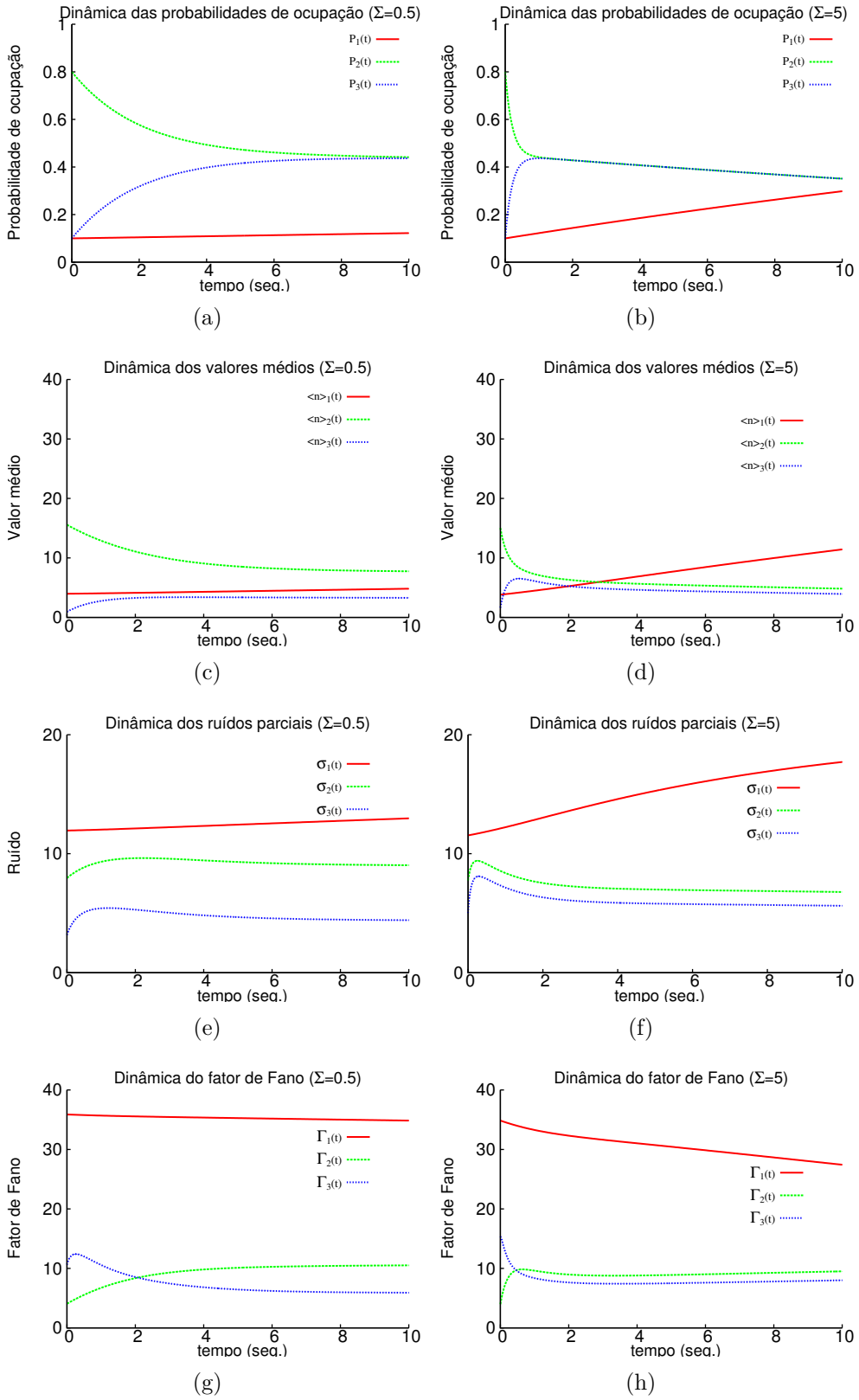

Figura 4.7 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 1/2$, $\epsilon_2 = 0.01$ e $\epsilon_3 = 0.99$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

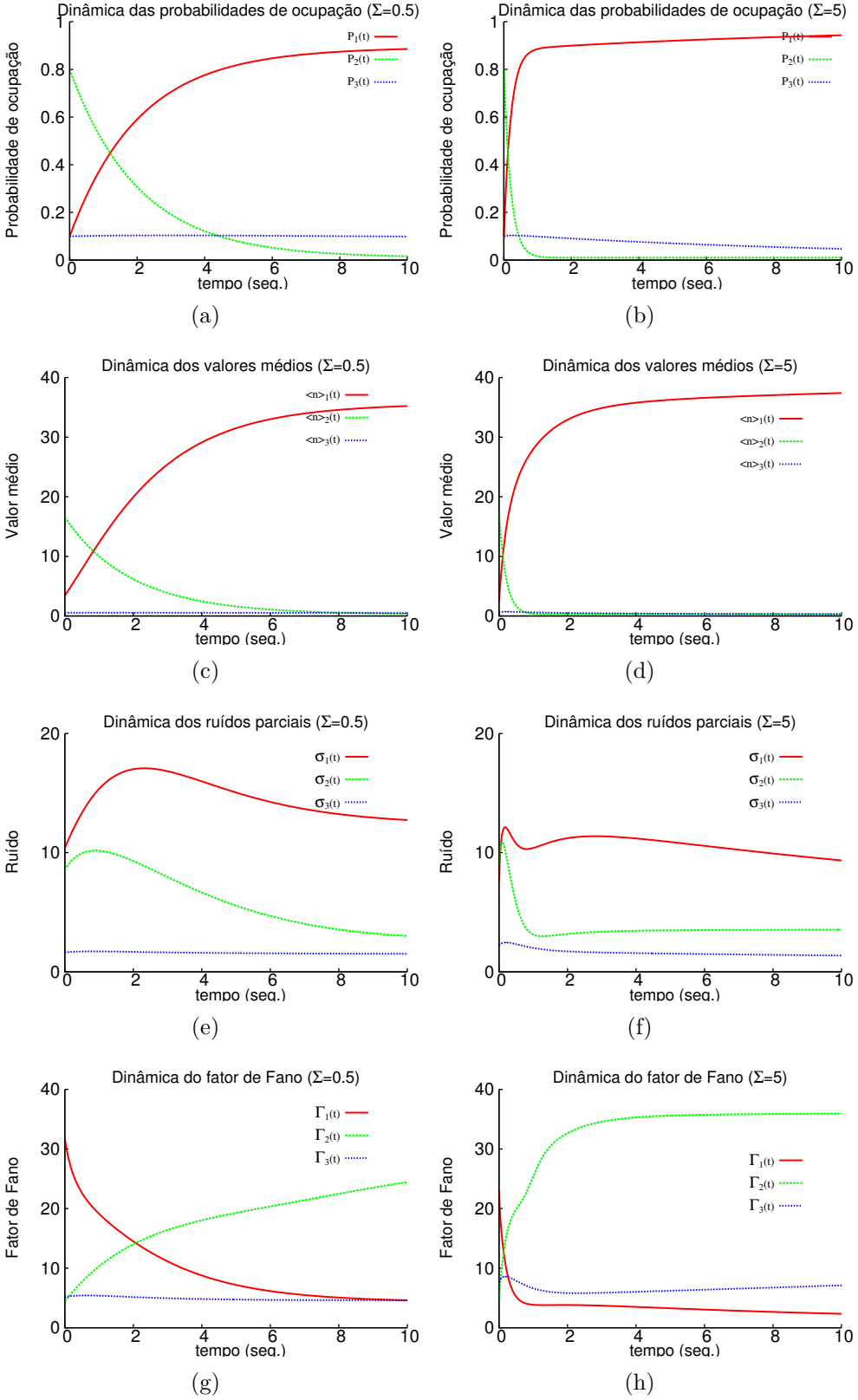

Figura 4.8 – Neste conjunto de gráficos mostramos o comportamento dinâmico dos momentos das distribuições, com ajuda da Eq.(4.19), no caso de switch rápido e lento. Aqui, o valor dos elementos da matriz ϵ são: $\epsilon_1 = 0.01$, $\epsilon_2 = 0.99$ e $\epsilon_3 = 1/2$. Os demais parâmetros têm os valores: $n_1 = 40$, $n_2 = 20$, $n_3 = 5$ e $\rho = 1$.

Capítulo 5

A generalização do modelo para um número arbitrário de estados

Neste capítulo iremos apresentar como o modelo pode ser generalizado para um número arbitrário de estados (N). Na primeira seção iremos mostrar como os parâmetros do modelo são expressos em termos das quantidades biológicas. Na seção seguinte mostraremos um exemplo de uma sistema que tem quatro estados de eficiência transcricional apresentando a matriz de acoplamento em termos dos parâmetros biológicos e usando a relação de recorrência para a expansão em momentos, exibimos as distribuições de probabilidade do sistema.

5.1 O espaço de parâmentros biológicos generalizado

Para começar nossa análise iremos partir da equação que governa a dinâmica das probabilidades de ocupação,

$$\frac{d\vec{P}(t)}{dt} = H\vec{P}(t) = \epsilon h\vec{P}(t), \tag{5.1}$$

onde $\vec{P}(t)$=$(P_1(t), ..., P_N(t))^T = \sum_{n=0}^{\infty} \vec{\phi}_n(t) = \vec{\phi}(1,t)$, onde N é um número inteiro e define a quantidade de estados do sistema. H é a matriz de acoplamento generalizada de

ordem $N \times N$, com suas propriedades dadas pela Eq.(2.7). Como discutido na seção 2 do capítulo 2 a escolha de uma base apropriada para se construir a matriz H conduz a fatoração da mesma em um produto de duas outras matrizes. Na Eq.(5.1) já usamos esta característica onde $H = \epsilon h$. A matriz ϵ é dada por:

$$\epsilon_{ii} = -1; \;\; 0 \leq \epsilon_{ij} \leq 1 \;\text{ se }\; i \neq j; \;\; \sum_{i=1}^{N} \epsilon_{ij} = 0, \tag{5.2}$$

e a matriz h é diagonal e seus elementos têm as propriedades

$$h_{ij} = h_i \;\; (h_i \in \Re^{+}) \;\text{ se }\; i = j; \;\; h_{ij} = 0 \;\text{ se }\; i \neq j. \tag{5.3}$$

Considerando a versão estacionária da Eq.(5.1) obtemos um sistema linear e homogêneo de equações que relaciona as probabilidades de ocupação assintóticas com os elementos da matriz de acoplamento,

$$H\vec{p} = \epsilon h \vec{p} = 0. \tag{5.4}$$

Introduzindo a matriz π, que é diagonal com elementos $\pi_{ii} = p_i$ $(i = 1, ..., N)$ e o vetor $\vec{h} = (h_1, ..., h_N)^T$, nós obtemos o sistema

$$\epsilon \pi \vec{h} = 0, \tag{5.5}$$

que é equivalente ao sistema da Eq.(5.4). Considerando o sistema homogêneo dado pela Eq.(5.5) nós podemos resolvê-lo para $N - 1$ componentes do vetor $\vec{h}$ e expressá-las como funções das probabilidades de ocupação assintóticas e dos elementos da matriz ϵ. Para se determinar todas as componentes do vetor $\vec{h}$ precisamos de mais uma equação que será dada pelo traço da matriz de acoplamento,

$$\mathrm{Tr}(H) = -\Sigma = -\sum_{i=1}^{N} h_i. \tag{5.6}$$

Com as Eqs.(5.5) e (5.6) estamos em posição de determinar todas as componentes do vetor $\vec{h}$ como função das probabilidades de ocupação assintóticas, Σ e os elementos da matriz ϵ, dando:

$$h_i = \frac{\Sigma}{\Delta} \frac{\delta_i}{p_i}, \tag{5.7}$$

onde δ_i são os cofatores dos elementos da diagonal da matriz ϵ, $\delta_i = \mathrm{Coef(i, i, \epsilon)}$, e

$$\Delta = \sum_{i=1}^{N} \frac{\delta_i}{p_i}. \tag{5.8}$$

Usando os elementos da matriz h como dados pela Eq.(5.7) notamos que a matriz de acoplamento terá a forma:

$$H = \frac{\Sigma}{\Delta} \epsilon \delta \pi^{-1}, \tag{5.9}$$

onde as matrizes ϵ e π já foram definidas e a matriz δ é a generalização daquela apresentada na seção 2 do capítulo 3, Eq.(3.30), é diagonal e seus elementos são os cofatores dos elementos da diagonal da matriz ϵ, ou seja, $\delta_{ii} = \mathrm{Coef(i, i, \epsilon)}$.

Com a matriz de acoplamento dada pela Eq.(5.9) e tendo como solução formal da Eq.(5.1),

$$\vec{P}(t) = \exp(Ht)\vec{P}(0) = \exp\left[\Sigma \left(\frac{\epsilon \delta \pi^{-1}}{\Delta}\right) t\right] \vec{P}(0), \tag{5.10}$$

podemos ver que Σ continua sendo a componente principal que governo a dinâmica do sistema, controlando se o switch é rápido ou lento, mesmo quando este pode transicionar entre um número arbitrário de estados com diferentes níveis de eficiência transcricional.

A matriz ϵ, e sua forma generalizada, tem todas as propriedades que sua parceira para $N = 3$ possui, assim como discutido na seção 2 do capítulo 2, e então, diversas estratégias de acoplamento entre os N estados do sistema são possíveis possibilitando uma maior flexibilidade do mesmo para se adaptar ao meio em que opera.

5.2 Distribuições de probabilidade para um sistema com quatro estados operacionais

Nesta seção exemplificamos os resultados da seção anterior para o caso em que o sistema em consideração pode operar em quatro estados com diferentes níveis de expressão.

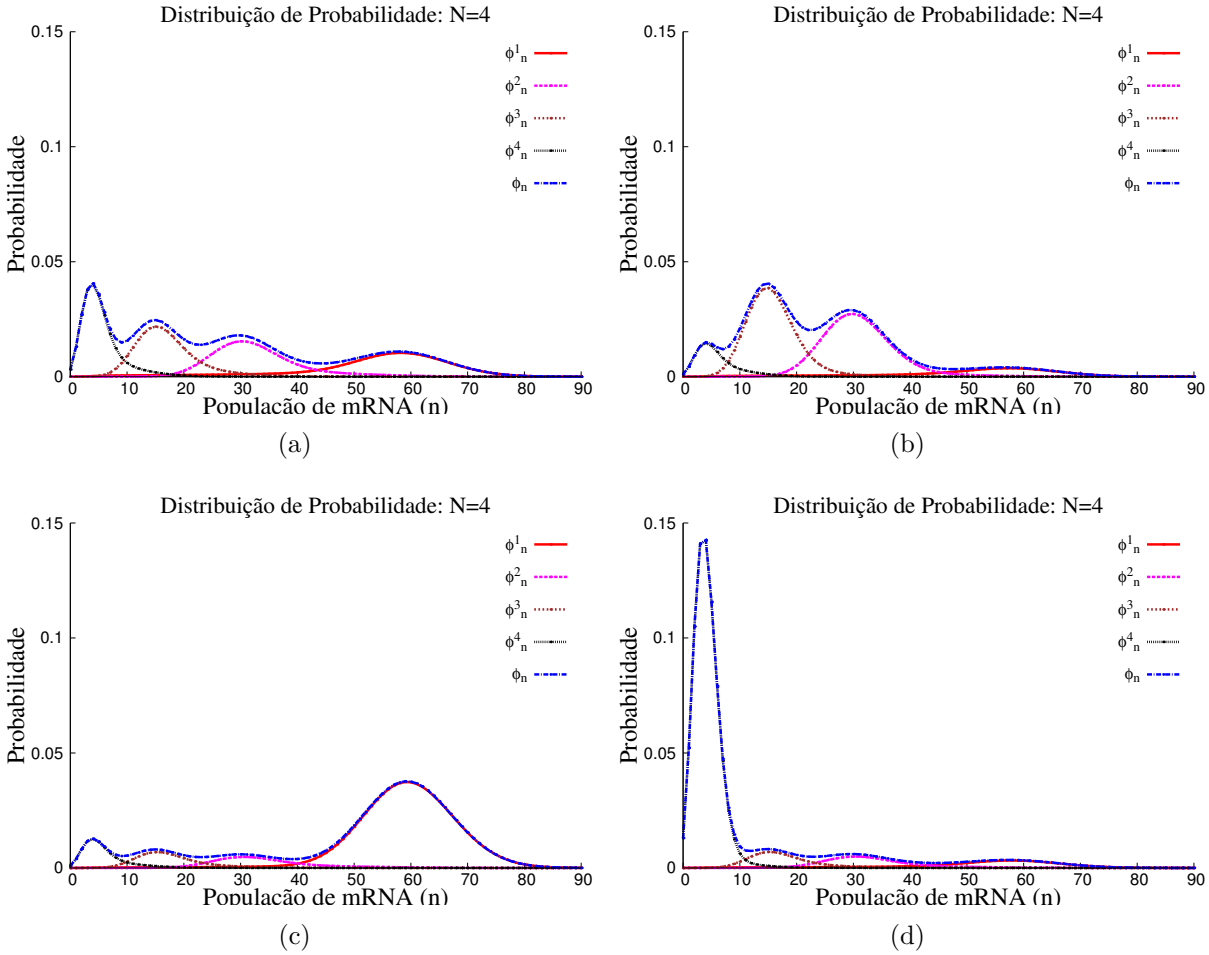

Figura 5.1 – Neste conjunto de gráficos mostramos o comportamento das distribuições de probabilidade de cada modo de eficiência transcricional assim como para a probabilidade total ($\phi_n = \phi_n^1 + \phi_n^2 + \phi_n^3 + \phi_n^4$) no regime de switch lento, $\Sigma = 0.5$. Na Fig.5.1(a) temos $p_1 = 1/4$, $p_2 = 1/4$, $p_3 = 1/4$ e $p_4 = 1/4$; Fig.3.1(b) temos $p_1 = 1/10$, $p_2 = 4/10$, $p_3 = 4/10$ e $p_4 = 1/10$; Fig.3.1(c) temos $p_1 = 3/4$, $p_2 = 1/12$, $p_3 = 1/12$ e $p_4 = 1/12$ e Fig.3.1(d) temos $p_1 = 1/12$, $p_2 = 1/12$, $p_3 = 1/12$ e $p_4 = 3/4$. Para todos os gráficos os valores dos demais parâmentros do modelo são: $n_1 = 60$, $n_2 = 30$, $n_3 = 15$ e $n_4 = 4$, $\epsilon_i = 0$ para $i = 1, ..., 8$ e $\rho = 1$.

Neste caso temos a matriz de eficiência do sistema como,

$$N = \begin{pmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_2 & 0 & 0 \\ 0 & 0 & n_3 & 0 \\ 0 & 0 & 0 & n_4 \end{pmatrix}, \tag{5.11}$$

A matriz de acoplamento, H, com ajuda das Eq.(5.7) à (5.9), é dada pelo produto de duas matrizes,

$$H = \frac{\Sigma}{\Delta}\epsilon h \tag{5.12}$$

onde a matriz ϵ contém a informação sobre o acoplamento entre os estados do sistema e

tem a forma:

$$\epsilon = \begin{pmatrix} -1 & \epsilon_3\epsilon_4 & \epsilon_5(1-\epsilon_6) & (1-\epsilon_7) \\ (1-\epsilon_1) & -1 & \epsilon_5\epsilon_6 & \epsilon_7(1-\epsilon_8) \\ \epsilon_1(1-\epsilon_2) & (1-\epsilon_3) & -1 & \epsilon_7\epsilon_8 \\ \epsilon_1\epsilon_2 & \epsilon_3(1-\epsilon_4) & (1-\epsilon_5) & -1 \end{pmatrix}. \tag{5.13}$$

A matriz h, que depende das condições do meio em que o sistema opera via as probabilidades de ocupação assintótica, assume a forma:

$$h = \begin{pmatrix} \delta_1 p_2 p_3 p_4 & 0 & 0 & 0 \\ 0 & p_1\delta_2 p_3 p_4 & 0 & 0 \\ 0 & 0 & p_1 p_2 \delta_3 p_4 & 0 \\ 0 & 0 & 0 & p_1 p_2 p_3 \delta 4 \end{pmatrix}. \tag{5.14}$$

Com as matrizes de eficiência transcriciona (N), ρ e a matriz de acoplamento podemos obter as distribuições de probabilidade que descrevem o sistema e também todos os momentos das distribuições. Como nosso intuito é apresentar a generalização da matriz de acoplamento com seus elementos expressos em termos das quantidades biológicas do modelo, vamos apenas calcular as distribuições de probabilidade usando a relação de recorrência,

$$\vec{\phi}_n = \frac{1}{n - H/\rho} N \vec{\phi}_{n-1}, \tag{5.15}$$

onde $\vec{\phi}_n = (\vec{\phi}_n^1, \vec{\phi}_n^2, \vec{\phi}_n^3, \vec{\phi}_n^4)^T$. Esta é a mesma relação de recorrência que foi utilizada no capítulo 3 para se obter as distribuições de probabilidade e os momentos das distribuições. Como lá, aqui precisamos ter conhecimento do coeficiente $\vec{\phi}_{n-1}$ para se calcular o coeficiente $\vec{\phi}_n$. Quando $n = 1$, $\vec{\phi}_0 = (p_1, p_2, p_3, p_4)^T$. Então, para se calcular todos os elementos da expansão em momentos 5.15 basta especificar o valor das probabilidades de ocupação assintóticas deste sistema de quatro estados, tendo em mente que: $p_1 + p_2 + p_3 + p_4 = 1$.

No conjunto de gráficos da Fig.(5.1) apresentamos quatro exemplos de configurações de equilíbrio para um sistema que possui quatro estados de eficiência transcricional e swithc lento, $\Sigma = 0.5$. Para todos os gráficos os elementos da matriz N são: $n_1 = 60$, $n_1 = 30$, $n_1 = 15$ e $n_1 = 4$. Para este exemplo, usamos o acoplamento cíclico, ou seja,

todos os elementos da matriz ϵ são iguais à zero ($\epsilon_i = 0$). Na Fig.5.1(a) as probabilidade de ocupação registram os valores: $p_1 = p_2 = p_3 = p_4 = 1/4$. Podemos observar uma estrutura multimodal com quatro picos indicando que o sistema pode operar em todos os seus estados. Na Fig.5.1(b) temos uma estrutura de dois picos centrais, em torno de $n = 15$ e $n = 30$ e outros dois picos residuais nos valores mais extremos do número da população de mRNA, $n = 60$ e $n = 4$, mostrando que o sistema opera predominantemente em seus estados de eficiência intermediária. Para este gráfico os valores das probabilidades de ocupação foram escolhidos como: $p_1 = 1/10, p_2 = 4/10, p_3 = 4/10$ e $p_4 = 1/10$. Nas Figs. 5.1(c) e 5.1(d) temos um pico dominante em torno de $n = 60$ e um em torno de $n = 4$, respectivamente. Na Fig.5.1(c) o gene produz predominantemente em seu estado de máxima eficiência com $p_1 = 3/4$ e na Fig.5.1(d) a produção de mRNA ocorre quase que exclusivamente no modo de eficiência transcricional mais baixa, com $p_3 = 3/4$.

Capítulo 6

Conclusões

Nesta tese propomos e resolvemos, exatamente, um modelo estocástico fenomenológico para o controle multimodal da expressão gênica em nível de transcrição. As soluções exatas foram calculadas usando a técnica das funções geradoras. Com estas soluções em mãos, obtivemos as distribuições de probabilidade multimodais e os momentos das distribuições. Para os momentos, foram obtidas as soluções exatas tanto no regime estacionário quanto dinâmico.

A definição de um espaço de parâmetros biológicos, que define e explicita o conteúdo biológico do modelo, só foi possível de ser obtido devido a escolha de uma base apropriada para a construção da matriz de acoplamento. Esta escolha mostrou que a matriz de acoplamento pode ser fatorizada e expressa como o produto de outras duas matrizes. A inspeção das matrizes revela que as propriedades do sistema também são fatorizadas e caracterizadas como de natureza intrínseca ou extrínseca. Na matriz ϵ estão codificadas propriedades intrínsecas do sistema sendo esta a responsável pela existência e a intensidade do acoplamento entre os estado. A segunda matriz, h, nesse espaço biológico, carrega em seus elementos, informações de natureza intrínseca do sistema e aquelas provenientes do meio ambiente. A parte intrínseca da matriz h é codificada no parâmetro Σ que mede a agilidade do switch de atingir a configuração de equilíbrio sendo então classificado como lento (Σ pequeno) ou rápido (Σ grande). A componente do meio é manifesta nas probabilidades de ocupação que irão medir se um estado do sistema pode ser utilizado para produção de mRNA ou não e com qual probabilidade.

Reescrevendo as equações diferenciais ordinárias do modelo em termos dos parâmetros do espaço biológico e introduzindo a variável w obtemos uma relação de recorrência para os coeficientes das funções geradoras que leva à solução exata do modelo apenas sendo necessário especificar as probabilidades de ocupação assintóticas de cada estado.

Com a solução exata em mãos, obtivemos as distribuições de probabilidade que mostram estruturas multimodais com N-picos ou estruturas com um pico predominante e outros residuais, dependendo da configuração de equilíbrio e da agilidade do sistema em alcançá-la. A vantagem da multimodalidade na expressão gênica, quando comparada com genes constitutivos, que têm seu nível médio de expressão fixo, ou com sistemas que admitem apenas o estados ligado ou desligado, vem do fato que a mera existência de moléculas reguladoras não é suficiente para que estas exercem seu papel regulatório, estrutural, de transporte, etc. Essas moléculas precisam, além de estarem presentes, acontecerem em concentrações ideais para que possam ser funcionais. Dois genes que são regulados pelo mesmo fator de transcrição podem ter afinidades distintas pelo mesmo. O gene com maior afinidade será ativado/reprimido em situações em que a concentração do agente regulador não atingiu um nível suficiente para regular o gene menos afim. Desta maneira, um gene constitutivo terá que produzir, no mínimo, um número tal de partículas que seja suficiente para atingir a concentração necessária para a regulação do gene com menor afinidade. No entanto, se o gene com maior afinidade for o mais vital para célula e seu parceiro menos afim poucas vezes solicitado para a manutenção da boa funcionalidade da mesma, a quantidade de partículas em excesso produzidas para se atingir a concentração ideal de regulação do gene menos requisitado, se realiza em desperdício de material e energia. A multimodalidade permite que este desperdício seja equacionado. Uma vez que o sistema pode operar em todos os seus estados (distribuições com N-picos) ou em um estado específico (distribuições com um pico predominante), este pode, com grande agilidade, e controlado pelas necessidades do meio, regular seu nível de expressão de forma a otimizar a produção de mRNA e assim evitando desperdícios.

A análise da estrutura de ruído mostrou que há duas componentes principais para redução do mesmo: a produção de mRNA ser feita em regime multimodal onde o estado

de eficiência intermediária é ocupado com probabilidade não desprezível ou o aumento na agilidade do gene de atingir sua configuração de equilíbrio, $\Sigma/\rho >> 1$. Que regimes de switch rápido levam a diminuição do ruído do sistema, já havia sido estabelecido em (49). No entanto, a redução via ocupação do estado intermediário é um novo mecanismo pela primeira vez observado e explorado. Esta nova componente de atenuação do ruído indica que a multimodalidade pode ter sido a estratégia escolhida pela natureza para se aumentar a precisão do processo transcricional quando necessário.

A seleção do esquema de acoplamento entre os estados do sistema revela: a existência de estados oscilantes fortemente amortecidos e mais uma fonte de flexibilidade do sistema que pode operar em regimes de maior ou menor ruído com diferentes níveis de expressão apenas selecionado os estados que farão parte do processo e como será a comunicação entre eles. Tal flexibilidade leva a um maior poder de adaptação por parte do sistema. A dinâmica das distribuições de probabilidade pode ser obtida exatamente de duas maneiras distintas. A primeira consiste em realizar uma transformação de variáveis, do tipo boost de Galileo, nas equações diferenciais parciais para as funções geradoras, que coloca as equações em suas versões estacionárias, como feito em (49). E a segunda, é usar os momentos dependentes do tempo das distribuições pois o conhecimento de todos os momentos (Eq.(4.19)), leva ao conhecimento integral das distribuições de probabilidade.

Em suma, temos um modelo estocástico, exatamente solúvel, para o controle da expressão gênica, em nível transcricional, que pode ser generalizado para incluir um número arbitrário de estados com diferentes níveis de eficiência para produção de mRNA. Tivemos como paradigma de produção multimodal o caso do ópero lac que opera com três estados distintos de eficiência transcricional. No entanto, a generalização do modelo, com um espaço de parâmetros biológico bem posto se faz necessário, quando temos interesse em analisar sistema de maior complexidade como, por exemplo, o λ-phage e a segmentação da Drosophila. Sistemas que possuem vários fatores de transcrição atuando simultaneamente e com várias regiões reguladoras envolvidas, são, inevitavelmente, multimodais.

REFERÊNCIAS

1 MORANGE, M. History of molecular biology. *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, 2009.

2 PTASHNE, M. *A genetic switch*: phage λ and hiher organisms. Cambridge: MA:Cell Press/Blacwell, 1992.

3 MONOD, J.; JACOB, F. Genetic regulatory mechanism in synthesis of protein. *Journal of Molecular Biology*, v. 3, n. 3, p. 318 - 356, 1961.

4 MULLER-HILL, B. *The lac operon*. A short history of a genetic paradigm. Berlin/New Yorl: Walter de Gruyter, 1996.

5 ALON, U. *A introduction to systems biology*: design principles of biological circuits. London: Chapman & Hall/CRC, 2007.

6 MAHESHRI, N.; O´SHEA, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annual Review of Biophysics and Biomolecular Structure*, v. 36, p. 413 - 434, 2007. DOI 10.1146/annurev.biophys.36.040306.132705

7 RAJ, A.; VAN OUDENAARDEN, A. Single-molecule approaches to stochastic gene expression. *Annual Review of Biophysics and Biomolecular Structure*, v. 38, p. 255 - 270, 2009. DOI 10.1146/annurev.biophys.37.032807.125928

8 ACKERS, G. K.; JOHNSON A. D.; SHEA, A. Quantitative model for gene regulation by λ phage repressor. *Proceedings of the National Academy of Sciences USA*, v. 79, n. 4, p. 1129 - 1133, 1982.

9 BECSKEI, A.; SERRANO, L. Engineering stability in gene networks by autoregulation. *Nature*, v. 405, n. 6786, p. 590 - 593, 2000.

10 BERG, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of Theoretical Biology*, v. 71, n. 4, p. 587 - 603, 1978.

11 BHALLA, U. S.; IYENGAR, R. Emergent properties of networks of biological signaling pathways. *Science*, v. 283, n. 5400, p. 381 - 387, 1999.

12 COOK, D. L.; GERBER, A. N.; TAPSCOTT, S. J. Modeling stochastic gene expression: implications for haploinsufficiency. *Proceedings of the National Academy of Sciences USA*, v. 95, n. 26, p. 15641 - 15646, 1998.

13 VON DASSOW, G. et al. The segment polarity network is a robust developmental module. *Nature*, v. 406, n. 6792, p. 188 - 192, 2000..

14 KENNELL, D.; RIEZMAN, H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. *Journal of Molecular Biology*, v. 144, n. 1, p. 1 - 21, 1977.

15 KO, M. S. H. A stochastic model for gene induction. *Journal of Theoretical Biology*, v. 153, n. 2, p. 181 - 194, 1991.

16 PUTTE, P.; GOOSEN, N. DNA inversions in phages and bacteria. *Trends in Genetics*, v. 8, n. 12, p. 457 - 462, 1992.

17 SIEGELE, D. A.; HU, J. C. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences USA*, v. 94, n. 15, p. 8168 - 8172, 1997.

18 DELBRUCK, M. Statistical fluctuations in autocatalytic reactions. *Journal of Chemical Physics*, v. 8, n. 1, p. 120 - 124, 1940.

19 GILLESPEI, D. T. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, v. 81, n. 25, p. 2340 - 2361, 1977.

20 WILKINSON, D. J. *Stochastic modelling for systems biology*. London: Chapman & Hall/CRC, 2006.

21 ELOWITZ, M.B. et al. Stochastic gene expression in a single cell. *Science*, v. 297, n. 5584, p. 1183 - 1186, 2002.

22 PAULSSON, J. Summing up the noise in gene networks. *Nature*, v. 427, n. 6973, p. 415 - 418, 2004.

23 McADAMS, H.H.; ARKIN, A. It´s a noisy business! genetic regulation at the nanomolar scale. *Trends Genetics*, v. 5, n. 2, p. 65 - 69, 1999.

24 McADAMS, H.H.; ARKIN, A. Gene regulation: towards a circuit engineering discipline. *Current Biology*, v. 10, n. 8, p. 814 - 819, 2000.

25 PAULSSON, J.; EHRENBERG, M. Stochastic focusing: random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Physical Review Letters*, v. 84, n. 23, p. 5447 - 5440, 2000.

26 BERG, O. G.; PAULSSON, J.; EHRENBERG, M. Fluctuations in repressor control: thermodynamic constraints on stochastic focusing. *Biophysical Journal*, v. 79, n. 6, p. 2944 - 2953, 2000.

27 METZLER, R.; WOLYNES, P.G. Number fluctuations and the threshold model of kinetic switches. *Chemical Physics*, v. 284, n. 1-2, p. 469 - 479, 2002.

28 WALCZAK, A.M. et al. Self-consistent proteomic field theory of stochastic gene switches. *Biophysical Journal*, v. 88, n. 2, p. 828-850, 2005.

29 PEDRAZZA J. M. et al. Noise propagation in gene networks. *Science*, v. 307, n. 5717, p. 1965 - 1969, 2005.

30 GARDNER, T.S. et al. Construction of genetic toggle switch in *Escherichia coli*. *Nature*, v. 403, n. 6767, p. 339 - 342, 2000.

31 HASTY, J. et al. Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences USA*, v. 97, n. 5, p. 2075 - 2080, 2000.

32 BLAKE, W.J. et al. Noise in eukaryotic gene expression. *Nature*, v. 422, n. 6932, p. 633 - 637, 2003.

33 PAULSSON, J. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proceedings of the National Academy of Sciences USA*, v. 97, n. 13, p. 7148 - 7153, 2000.

34 ELOWITZ, M.B.; LEIBLER, S. A synthetic oscillatory network of transcriptional regulators. *Nature*, v. 403, n. 6767, p. 335 - 338, 2000.

35 McADAMS, H.H.; ARKIN, A. Stochastic mechanisms in gene expression *Proceedings of the National Academy of Sciences USA*, v. 94, n. 3, p. 814 - 819, 1997.

36 McADAMS, H.H.; ARKIN, A. Simulation of prokaryotic genetic circuits. *Annual Review of Biophysics and Biomolecular Structure*, v. 27, p. 199 - 224, 1998. DOI 10.1146/annurev.biophys.27.1.199.

37 THATTAI, M.; VAN OUDENAARDEN, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences USA*, v. 98, n. 15, p. 8614 - 8619, 2001.

38 OZBUDAK, E.M. et al. Regulation of noise in the expression of single gene. *Nature Genetics*, v. 31, n. 1, p. 69 - 73, 2002.

39 KEPLER T. B.; ELSTON, T. C. Stochasticity in transcriptional regulation: origins, consequences and mathematical representations *Biophysical Journal*, v. 81, n. 6, p. 3116 - 3136, 2001.

40 SASAI, M.; WOLYNES, P.G. Stochastic gene expression as a many-body problem. *Proceedings of the National Academy of Sciences USA*, v. 100, n. 5, p. 2374 - 2379, 2003.

41 HORNOS, J.E.M. et al. Self-regulating gene: an exact solution. *Physical Review E*, v. 72, n. 5, p. 051907, 2005.

42 INNOCENTINI, G. C. P.; HORNOS, J. E. M. Modeling stochastic gene expression under repression. *Journal of Mathematical Biology*, v. 55, n. 3, p. 413 - 431, 2007.

43 RAMOS, A. F.; INNOCENTINI, G. C. P.; HORNOS, J. E. M. Exact time-dependent solutions for a self-regulating gene. *Physical Review E*, v. 83, n. 6, p. 062902, 2011.

44 RAMOS, A. F. et. al. Symmetry in biology: from genetic code to stochastic gene regulation. *IET Systems Biology*, v. 4, n. 5, p. 311 - 329, 2010.

45 RAMOS, A. F.; HORNOS, J. E. M. Symmetry and stochastic gene regulation. *Physical Review Letters*, v. 99, n. 10, p. 108103, 2007.

46 VAN KAMPEN, N. G. *Stochatic processes in physics and chemistry.* Amsterdam: North-Holland, 1992.

47 COX, D. R.; MILLER, H. D. *The theory of stochastic processes.* London: Chapman and Hall, 1965.

48 NOVICK, A.; WEINER, M. Enzyme induction as an all-or-nonephenomenon. *Proceedings of the National Academy of Sciences USA*, v. 43, n. 7, p. 553 - 566, 1957.

49 INNOCENTINI, G. C. P. *Modelamento estocástico para a expressão gênica.* 2008. 53 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.