

## Path integral formulation and Feynman rules for phylogenetic branching models

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2005 J. Phys. A: Math. Gen. 38 9621

(<http://iopscience.iop.org/0305-4470/38/44/002>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.94

The article was downloaded on 03/06/2010 at 04:01

Please note that [terms and conditions apply](#).

## Path integral formulation and Feynman rules for phylogenetic branching models

**P D Jarvis, J D Bashford and J G Sumner**

School of Mathematics and Physics, University of Tasmania, GPO Box 252C, 7001 Hobart, TAS, Australia

E-mail: [Peter.Jarvis@utas.edu.au](mailto:Peter.Jarvis@utas.edu.au), [James.Bashford@utas.edu.au](mailto:James.Bashford@utas.edu.au)  
and [Jeremy.Sumner@utas.edu.au](mailto:Jeremy.Sumner@utas.edu.au)

Received 1 December 2004

Published 19 October 2005

Online at [stacks.iop.org/JPhysA/38/9621](http://stacks.iop.org/JPhysA/38/9621)

### Abstract

A dynamical picture of phylogenetic evolution is given in terms of Markov models on a state space, comprising joint probability distributions for character types of taxonomic classes. Phylogenetic branching is a process which augments the number of taxa under consideration, and hence the rank of the underlying joint probability state tensor. We point out the combinatorial necessity for a second-quantized, or Fock space setting, incorporating discrete counting labels for taxa and character types, to allow for a description in the number basis. Rate operators describing both time evolution without branching, and also phylogenetic branching events, are identified. A detailed development of these ideas is given, using standard transcriptions from the microscopic formulation of non-equilibrium reaction–diffusion or birth–death processes. These give the relations between stochastic rate matrices, the matrix elements of the corresponding evolution operators representing them, and the integral kernels needed to implement these as path integrals. The ‘free’ theory (without branching) is solved, and the correct trilinear ‘interaction’ terms (representing branching events) are presented. The full model is developed in perturbation theory via the derivation of explicit Feynman rules which establish that the probabilities (pattern frequencies of leaf colourations) arising as matrix elements of the time evolution operator are identical with those computed via the standard analysis. Simple examples (phylogenetic trees with two or three leaves), are discussed in detail. Further implications for the work are briefly considered including the role of time reparametrization covariance.

PACS numbers: 02.50.Ga, 31.15.Kb

Mathematics Subject Classification: 05C05, 60J80, 92D20, 81S40

## 1. Introduction and background

The use of Markov models of change to taxonomic character probability distributions is a standard technique for describing mutations, and for inferring ancestral relationships between taxa. A general stochastic framework for phylogenetic branching models is as follows. By assumption, different ‘taxonomic units’ are identified, and classified by a set of defining characteristics: based on morphological features for example, or on sequence data for a particular gene or protein say. To each taxon is ascribed a probability density on the set of characters, and it is the task of phylogenetic reconstruction to infer ancestral relationships within a group of taxa, given observed pattern frequencies for characters amongst the taxa (definitions are given later in the text). In such phylogenetic reconstruction, the Markov chain model describing the stochastic evolution of characters is extended appropriately to encompass ‘branching’ where the number of taxa is augmented as new taxonomic types evolve (for example, by speciation or gene duplication), from an initial single progenitor, through to the final number of types under study. For details of the subject, including overviews of applications, current problems and new directions, we refer to recent textbooks, for example [9, 21].

In recent work [2, 10, 23, 24] it has been pointed out that a fruitful approach to phylogenetic analysis is afforded by taking the formal perspective of multilinear tensor algebra familiar from physical systems. For example, in the analysis of symmetry properties (of the Markov rate matrix, and of the branching process) it is natural to consider *continuous* Lie transformation groups acting on the tensor spaces, and the associated representation theory [23]. Furthermore, a remarkable analogy between branching processes (where the technical constraint of local conditional independence [21] is imposed) and state entanglement in quantum physics has been noted [24]. In particular, for two characters (equivalent to single qubit (2-state) systems in quantum physics) the well-known log det distance measure for two taxa is essentially the *concurrence* (for two qubits, related to the von Neumann entropy of a partial density operator); equally the *tangle* (an entanglement measure for three qubits) has been proposed as a useful measure of distance for three taxa in the two-character case, and the analysis of its properties in the phylogenetic context has been initiated [24, 25]. Further applications of classical invariant theory for phylogenetic analysis are developed in [10].

In the letter [2] it was argued that a further useful perspective on phylogenetics, again inspired by physics, can be gained by interpreting ‘branching’ in the model as a linear operator which *augments* the rank of the tensor corresponding to the joint probability distribution of character types (see also [23, 24]). In order to regard the entire model, including especially the time development represented by the branching dynamics, in a uniform way, it is natural to seek a setting in multilinear algebra where the linear space describing state probabilities for taxa can be lifted to an appropriate free algebra in the sense of tensor products, or ‘Fock space’ in physical language, so that the linear ‘branching operator’ has a uniform (extended) domain of definition. Possible interaction terms representing this operator, corresponding to phylogenetic branching events, can then readily be implemented in the language of second quantization as shown in [2]. Although formal, the transcription to physical language provided does indeed establish that the entire Markov branching model can be regarded as a standard Markov chain, but with dynamics on a suitably extended state space—a fact not noted explicitly before. With closed form expressions for the probabilities in hand, it may also be possible to investigate these from various analytical viewpoints not accessible hitherto. Moreover, the physical language is quite flexible, and may suggest useful insights into the models as well as generalizations.

In the present paper, a further step towards such analytical investigations and generalizations is taken, in that the second-quantized framework is transcribed into the language of path integrals. The dynamical quantities of interest become phylogenetic ‘path’ variables (or ‘classical fields’), defined over a discrete spatial lattice. Time evolution of the system is developed in perturbation theory, yielding standard probabilities as convolutions of the appropriate kernel with the initial probability distribution, that is, as matrix elements of the evolution operator. Similar models of reaction–diffusion or birth–death processes have been extensively investigated [5, 6, 13, 18] so that there is a wealth of technical experience within this approach, and possibilities for generalization. These introductory comments are supplemented in the conclusions by further discussion of possible applications (see summary below).

The outline of the paper is as follows<sup>1</sup>. In section 2, we give an analysis of standard accounts of phylogenetic processes (as used, for example, in analyses for inferring ancestral trees) to justify our claim that a multilinear tensor description is appropriate, and equivalent to the usual approach. A standard notation is introduced including the branching or ‘splitting’ operator whose properties are discussed. In section 3 the rate operator and the branching operator are reformulated as interaction terms in an extended time evolution over Fock space. Attention is given to the ‘copy space’ needed to identify taxa—both for the observed taxa (leaves) and ancestral stages (‘internal’ edges of the phylogenetic tree)—and it is argued that for models with  $L$  distinguished leaves, a  $2^L$ -dimensional ‘label’ space is needed. Label summations suggest a natural identification with the ‘momentum’ space for periodic functions over a hypercubic ‘spatial’ lattice in  $L$  dimensions (with  $2^L$  nodes in the unit cell), leading to the possibility of viewing the system dually in ‘position’ space. Section 4 gives a brief pedagogical review of standard path integral techniques as applied for the analysis of non-equilibrium reaction–diffusion systems in a microscopic approach. In section 5 these ingredients are synthesized in a path integral formulation for a ‘free’ phylogenetic system, that is, a collection of up to  $L$  taxa with no phylogenetic association (not necessarily in a stationary state). It is shown that the abstract dynamics, represented by the evolution kernel of the system in the path integral approach which is formulated and derived explicitly, does indeed make the system evolve in a standard way according to a continuous Markov branching process. In section 6 the question of the branching operator is resumed, and plausible interaction terms (and corresponding normal kernels) are introduced in the path integral language. It is shown in simple examples (trees with two or three leaves) that, in both the operator and path integral language picture, the probabilities arising as matrix elements from the dynamics of the model are identical to those computed in standard likelihood analyses for inferring phylogenetic trees. This is borne out in the appendix where formal Feynman rules are derived directly from perturbation theory, and which can immediately be seen to encode the usual sum over extended leaf colourations presentations. The conclusions, section 7, reiterate the main points of the paper and further implications and applications of our work are briefly discussed. In particular, we comment on the role of the group of time reparametrizations (diffeomorphisms), in the issue of assigning ‘true’ historical time to phylogenetic events.

## 2. Tensor methods and stochastic models of phylogenetic branching

It is usual to pose the standard stochastic model of phylogenetics by stating transition probabilities [3, 9, 12, 21]. It is, however, possible to present the same system in an abstract

<sup>1</sup> For the benefit of readers unfamiliar with the subject matter, technicalities in various sections below are treated as fully as possible.

multilinear tensor setting. Our philosophy here is similar to that of [17] (see also [4]). In such a formulation, the evolution of the phylogenetic system is represented by a group action on a tensor product space, with the branching structure formalized by the introduction of linear ‘splitting’ operators which increment the rank of the tensor space. As pointed out in the introduction, this basis-independent description has many advantages, prompting the investigation of the rich algebraic structure of the system. The door is opened to the discussion of symmetry groups and subgroups, representation theory and diagonalization, the differential structure of the rate operators and orbit classes of their action, and the ring structure of invariant functions (see [10, 23, 24]).

Introduce a set,  $\mathcal{K}$ , which consists of  $K$  discrete elements, conventionally labelled by the integers  $\{0, 1, 2, \dots, K - 1\}$ . Consider a system consisting of  $N$  ‘samples’ to each of which can be attributed one of  $K$  distinct characters. Associated with such a system we have the set of frequencies

$$\widehat{p}^\alpha := \frac{\text{total number of occurrences of character } \alpha}{N}, \quad \alpha = 0, 1, \dots, K - 1.$$

In particular, we are interested in the character frequencies occurring in the genome of a given taxon. The archetypical example is that of the DNA sequence, where the ‘samples’ are sites, and with four characters  $\{A, G, C, T\}$ , but it is, of course, possible to envisage the use of other character sets derived from the molecular data, so  $K$  is left general in this discussion. Examples include the amino acids ( $K = 20$ ), codons ( $K = 64$ ) or instead of nucleic acid bases themselves, a binary pyrimidine/purine  $Y/R$  classification of them ( $K = 2$ ). For practical purposes, the usual practice is to take one particular gene of an organism as being the representative for the taxon class, although it would be possible to sample a whole genome or set of genomes and calculate the character frequencies across that set, and take those frequencies as the representative of that taxon. Practical considerations aside, we proceed to model the time evolution of these frequencies stochastically.

Introduce a random variable  $X$  which takes on values in  $\mathcal{K}$ . It is necessary to define a set of time-dependent probabilities which are the theoretical limit

$$p^\alpha(t) := \mathbb{P}(X = \alpha, t) = \lim_{N \rightarrow \infty} \widehat{p}^\alpha(t). \quad (2.1)$$

The stochastic evolution of the probabilities is assumed to satisfy the continuous time Markov property, that is, the state at time  $t$  depends only on the immediately preceding state at time  $t - \delta t$  say, and hence

$$p^\alpha(t) = \sum_{\beta \in \mathcal{K}} \mathbb{P}(X = \alpha, t | X = \beta, t - \delta t) p^\beta(t - \delta t), \quad (2.2)$$

which in turn implies, assuming linearity and differentiability, that

$$\frac{d}{dt} p^\alpha(t) = \sum_{\beta \in \mathcal{K}} \lim_{\delta t \rightarrow 0} \frac{\mathbb{P}(X = \alpha, t + \delta t | X = \beta, t) - \delta^\alpha_\beta}{\delta t} p^\beta(t). \quad (2.3)$$

We define the (time-dependent) rate matrix

$$R^\alpha_\beta(t) = \lim_{\delta t \rightarrow 0} \frac{\mathbb{P}(X = \alpha, t + \delta t | X = \beta, t) - \delta^\alpha_\beta}{\delta t}. \quad (2.4)$$

To preserve reality of the probabilities and the property  $\sum_\alpha p^\alpha(t) = 1$  for all  $t$  it follows that  $R$  is a real-valued zero column sum matrix. In order to preserve positivity of the probabilities it must also be the case that for all  $t$

$$R^\alpha_\beta(t) \geq 0, \quad \forall \alpha \neq \beta, \quad R^\alpha_\alpha(t) \leq 0 \quad (\text{no sum}). \quad (2.5)$$

For a homogeneous model, the rate matrix is assumed to be time *independent*, with solution

$$p^\alpha(t) = \sum_{\beta} M^\alpha_{\beta}(t) p^\beta(0), \quad M^\alpha_{\beta}(t) = [e^{Rt}]^\alpha_{\beta}, \tag{2.6}$$

where  $\exp(Rt)$  is calculated using matrix multiplication.

Phylogenetics is concerned with deriving the past evolutionary relationships of multiple taxa. As already mentioned, the modern approach is to compare the genomes of the taxa. An essential part of the analysis is the ability to align the genomes of distinct taxa successfully. (The possibility or otherwise of such alignment is not discussed here.) Having aligned the genomes it is possible to calculate *pattern* frequencies. These patterns are read off ‘vertically’ across the aligned sequences. The data are then

$$\widehat{P}^{\alpha_1\alpha_2\cdots\alpha_L} := \frac{\text{total number of occurrences of pattern } \alpha_1\alpha_2\cdots\alpha_L}{N},$$

$$\alpha_1, \alpha_2, \dots, \alpha_L = 0, \dots, K - 1.$$

Introduce random variables  $X_1, X_2, \dots, X_L$  each of which takes on values in the individual character spaces  $\mathcal{K}$ , and  $X = (X_1 X_2 \cdots X_L)$  which takes on values in the  $L$ -component character space  $\mathcal{K} \times \mathcal{K} \cdots \times \mathcal{K}$ . We model these pattern frequencies by again defining a set of time-dependent probabilities which are the theoretical limit

$$P^{\alpha_1\alpha_2\cdots\alpha_L}(t) := \mathbb{P}(X = \alpha_1\alpha_2\cdots\alpha_L, t) = \lim_{N \rightarrow \infty} \widehat{P}^{\alpha_1\alpha_2\cdots\alpha_L}(t).$$

The Markov property for this system is expressed as the dependence of  $\mathbb{P}(X = \alpha_1\alpha_2\cdots\alpha_L, t)$  only on its values at the immediately preceding time,  $t - \delta t$  say. It is also assumed that the transition probabilities are conditionally independent across different taxa. This is a standard assumption [9, 12, 21] and is quite well founded from a biological perspective. Again assuming differentiability and linearity, a solution is found to be

$$P^{\alpha_1\alpha_2\cdots\alpha_L}(t) = \sum_{\beta_1, \beta_2, \dots, \beta_L} M_1^{\alpha_1}_{\beta_1}(t) M_2^{\alpha_2}_{\beta_2}(t) \cdots M_L^{\alpha_L}_{\beta_L}(t) P^{\beta_1\beta_2\cdots\beta_L}(0). \tag{2.7}$$

The final part of the model is to introduce the branching. In the case of two taxa diverging from a common ancestor, considering that at the time of branching the character frequencies are identical, the correct formula for the pattern frequencies is given by (see, for example, [16])

$$P^{\alpha_1\alpha_2}(t) = \sum_{\beta \in \mathcal{K}} M_1^{\alpha_1}_{\beta}(t) M_2^{\alpha_2}_{\beta}(t) p^\beta(0), \tag{2.8}$$

as will be derived in detail below. This situation can then be iterated for the case of arbitrary trees (see, for example, [7, 8] as well as the standard texts already cited). Having given the standard stochastic model of phylogenetics we proceed to abstract the presentation. Introduce the vector space<sup>2</sup>  $V \cong \mathbb{C}^K$ , with preferred basis  $\{e_0, e_1, \dots, e_{K-1}\}$ . We associate the set of probabilities (2.1) with the unique vector

$$p^\alpha(t) \rightarrow p(t) = \sum_{\alpha \in \mathcal{K}} p^\alpha(t) e_\alpha. \tag{2.9}$$

Having made this abstraction it is possible to view the stochastic evolution given by (2.6) as linear group action on  $V$ , clearly an appropriate 1-parameter subgroup of  $GL(K)$ . The structure of the Markov group is discussed in [14], and from the viewpoint of invariant theory

<sup>2</sup> Although the above presentation involves only *real* numbers, we work over  $\mathbb{C}$  to allow for the use of more convenient symmetry adapted bases, or other ways of diagonalizing rate matrices [23]. Of course, measurable quantities are referred back to the distinguished basis at the end of the analysis.

in [10, 24, 25]. For the case of phylogenetics, the obvious generalization is to the tensor product space  $V^{\otimes L}$ , with group action as the appropriate subgroup of the direct product group  $GL(K)^{\times L}$ . The final step is to describe the branching process upon this tensor product space.

In order to formalize this we introduce the *splitting* operator  $\delta : V \rightarrow V \otimes V$ . Progress is made by simply expressing the most general action of  $\delta$  on the basis elements of  $V$ ,

$$\delta \cdot e_\alpha = \sum_{\alpha, \beta, \gamma} \Gamma_\alpha^{\beta\gamma} e_\beta \otimes e_\gamma, \tag{2.10}$$

where  $\Gamma_\alpha^{\beta\gamma}$  are an arbitrary set of coefficients. Imposing conditional independence upon the distinct branches in order to constrain these coefficients, we need only consider initial probabilities of the form

$$p_{(\lambda)}^\alpha = \delta_\lambda^\alpha, \quad \gamma = 0, 1, \dots, K - 1. \tag{2.11}$$

Consider a branching even at time  $t$  so that the initial single taxon state a small time before branching is

$$p_{(\lambda)}(t) = \sum_\alpha p_{(\lambda)}^\alpha(t) e_\alpha = \sum_\alpha \delta_\lambda^\alpha e_\alpha = e_\lambda. \tag{2.12}$$

Directly subsequent to the branching event the 2-taxon state is therefore given by

$$P_{(\lambda)}(t) = \delta \cdot p_{(\lambda)}(t) = \sum_{\alpha, \beta, \gamma} \delta_\lambda^\alpha \Gamma_\sigma^{\rho\rho'} e_\rho \otimes e_{\rho'}. \tag{2.13}$$

On the other hand, conditional independence leads to

$$\begin{aligned} \mathbb{P}(X = \alpha_1 \alpha_2, t + \delta t | X_1 = X_2 = \lambda, t) \\ = \mathbb{P}(X_1 = \alpha_1, t + \delta t | X_1 = \lambda, t) \cdot \mathbb{P}(X_2 = \alpha_2, t + \delta t | X_2 = \lambda, t). \end{aligned} \tag{2.14}$$

Using the tensor formalism these transition probabilities can be expressed separately as

$$\begin{aligned} \mathbb{P}(X_1 = \alpha_1, t + \delta t | X_1 = \lambda, t) &= \sum_\rho M_1^{\alpha_1 \rho}(\delta t) p_{(\lambda)}^\rho(t), \\ \mathbb{P}(X_2 = \alpha_2, t + \delta t | X_2 = \lambda, t) &= \sum_{\rho'} M_2^{\alpha_2 \rho'}(\delta t) p_{(\lambda)}^{\rho'}(t). \end{aligned} \tag{2.15}$$

However, from (2.7) we have

$$\begin{aligned} \mathbb{P}(X = \alpha_1 \alpha_2, t + \delta t | X_1 = X_2 = \lambda, t) \\ = \sum_{\rho, \rho', \sigma} M_1^{\alpha_1 \rho}(\delta t) M_2^{\alpha_2 \rho'}(\delta t) \delta_\lambda^\sigma \Gamma_\sigma^{\rho\rho'}. \end{aligned}$$

Implementing (2.14), (2.12) and considering the limit  $\delta t \rightarrow 0$  with  $M^\alpha_\rho(\delta t) \rightarrow \delta^\alpha_\rho$  then leads to the requirement that

$$\Gamma_\lambda^{\rho\rho'} = \delta_\lambda^\rho \delta_\lambda^{\rho'}, \tag{2.16}$$

and the definition for the splitting operator in the preferred basis becomes simply

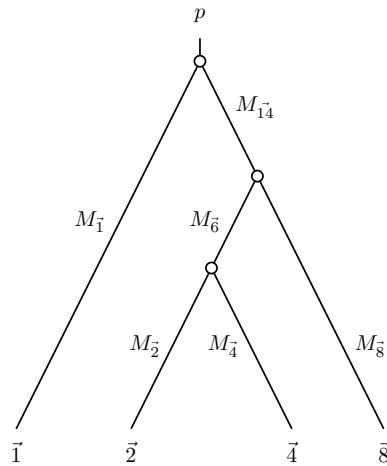
$$\delta \cdot e_\alpha = e_\alpha \otimes e_\alpha. \tag{2.17}$$

Using the above notation, we are now in a position to write down formally expressions for the probabilities on arbitrary trees. As an example, the expression which defines the general Markov model on the tree  $(1((23)4))$  (or<sup>3</sup>  $(\bar{1}((\bar{2}\bar{4})\bar{3}))$ ) is given by (see figure 1)

$$P_{(\bar{1}((\bar{2}\bar{4})\bar{3}))} = (M_{\bar{1}} \otimes M_{\bar{2}} \otimes M_{\bar{4}} \otimes M_{\bar{3}}) 1 \otimes \delta \otimes 1(1 \otimes M_{\bar{6}} \otimes 1) 1 \otimes \delta(1 \otimes M_{\bar{14}}) \delta \cdot p \tag{2.18}$$

where  $p$  is the initial single taxon distribution and the  $M$  are arbitrary stochastic operators (Markov matrices) on the designated edges.

<sup>3</sup> In terms of the binary labelling introduced below, this tree is  $(\bar{1}((\bar{2}\bar{4})\bar{3}))$ , with the remaining edge assignments determined additively.



**Figure 1.** The general Markov model for four taxa with tree  $1((23)4)$  (or  $\overline{1}((\overline{2}\overline{4})\overline{8})$ ) in terms of binary edge labelling. The  $M$  are arbitrary transition probabilities (Markov matrices) on the designated edges.

### 3. Fock space and momentum labels for binary trees

In the previous section, we have presented a description of phylogenetic systems in terms of a multilinear tensor calculus based on copies of the basic state space  $V \simeq \mathbb{C}^K$ . This comprises vectors with positive coefficients  $p^\alpha$  in the distinguished basis, corresponding to the theoretical probabilities for observation of a particular character  $\alpha, \alpha = 0, \dots, K - 1$ ; higher rank tensors  $P^{\alpha_1\alpha_2\dots\alpha_n}$  represent joint probability densities. Moreover, we introduced a linear operator  $\delta : V \rightarrow V \otimes V$ , again defined by its matrix elements in the distinguished basis, representing phylogenetic branching viewed dynamically as an event occurring at a specific time.

In this and the following sections, we wish to argue for a more universal view which is appropriate for *arbitrary* trees. Given that branching might occur at various times, this means that the ‘state space’ might be anything from  $V$  (for the root edge of the tree), to  $V \otimes V$  (if there is only one branching), and so on, up to  $V \otimes V \otimes V \otimes \dots \otimes V, L$  times, if the final number of taxa (the number of leaves of the tree) is  $L$ . The only logical way to encompass all these possibilities within one description in linear algebra is to adopt as the proper state space, an appropriate *Fock space*  $F$  associated with  $V$ , in this case, for example,

$$\begin{aligned}
 F^L &= \mathbb{C} \oplus V \oplus V \otimes V \oplus \dots \oplus V \otimes V \otimes \dots \otimes V \\
 &= \bigoplus_{n=0}^L (\otimes^n V).
 \end{aligned}
 \tag{3.19}$$

The advantage of this formal change of perspective is that it allows both the normal time evolution (as described above, the Markov rate operator acting on each copy of  $V$ ) and the branching operator (as described above)  $\delta$ , or its natural extensions  $\mathbb{1} \otimes \mathbb{1} \otimes \dots \otimes \delta \otimes \dots \otimes \mathbb{1}$ , simultaneously to be regarded as operators on  $F$ .

In physical settings, it is conventional to apply the above construction for the description of ‘composite’ systems where the state space  $V$  corresponds to a single subsystem, and the tensor products allow for copies of  $V$  corresponding to different numbers of subsystems. In relativistic systems, this is of course the setting for elementary particle interactions, but the same idea is also appropriate in the non-relativistic case. However, in quantum systems the



Pauli principle mandates that the general spaces  $V \otimes V \otimes \dots \otimes V$  are too big—the individual subsystems are *indistinguishable* in that the ordering of individual state vectors in the tensor product is immaterial (up to a possible sign factor for fermionic systems). This means that the relevant Fock spaces are technically speaking the linear spaces  $F^+$ ,  $F^-$  associated, respectively, with the symmetric (for bosons), or (for fermions) the antisymmetric or exterior tensor algebras:

$$\begin{aligned} F^+ &= \mathbb{C} \oplus V \oplus V \vee V \oplus \dots \oplus V \vee V \vee \dots \vee V \dots, \\ &= \bigoplus_{n=0}^{\infty} (\vee^n V), \\ F^- &= \mathbb{C} \oplus V \oplus V \wedge V \oplus \dots \oplus V \wedge V \wedge \dots \wedge V, \\ &= \bigoplus_{n=0}^L (\wedge^n V) \end{aligned} \quad (3.20)$$

In adopting the machinery of Fock space to the phylogenetic context, the ‘subsystems’ become the individual taxa extant at any particular stage of the branching process, and the (anti)symmetrization principle would need to be interpreted as saying that all taxa are equivalent, or that the tensor probability density of rank  $n$  is totally symmetric or totally antisymmetric. Thus for a given choice of observed characters represented by the *symmetric* probability density  $P^{\alpha_1 \alpha_2 \dots \alpha_n}$ , it would be immaterial which taxon (from 1 to  $n$  in this case) carried which character,

$$P^{\alpha_1 \alpha_2 \dots \alpha_n} = P^{\alpha_{\sigma 1} \alpha_{\sigma 2} \dots \alpha_{\sigma n}} \quad (3.21)$$

for any permutations  $\sigma$ . In phylogenetic branching, this symmetrization may well be appropriate for cases where it is suspected that a number of siblings are diverging from a common origin with *equal* rate matrices<sup>4</sup>, but in general, we would certainly wish to be able to distinguish between taxa.

These considerations imply that the higher rank tensor spaces  $V \otimes V \otimes \dots \otimes V$  introduced above should be regarded technically as products of a number of *labelled* spaces, for example for the final  $L$  taxon system,  $V_1 \otimes V_2 \otimes \dots \otimes V_L$  where each  $V_n$  is a *distinct* copy of  $V$ ,  $V_n \simeq V$ ,  $n = 1, \dots, L$ . However, since the  $n$  taxon spaces required for the system at earlier times (arising from branching at intermediate nodes above the leaves of the tree) can comprise *any* subsets of the labels  $1, \dots, L$ , we are led necessarily to a labelling system appropriate to the power set  $2^L$ , or simply to the well-known system of edge labelling for binary trees, by binary  $L$ -vectors, whereby leaf edges are labelled by powers or decimal equivalents  $1, 2^1, 2^2, \dots, 2^{L-1}$ , and the assignments for the remaining edges determined additively (for an example, see figure 2).

To this end we therefore introduce the following (extended) Fock space (we discuss only the bosonic case in this paper):

$$\begin{aligned} \mathcal{F}^+ &= \mathbb{C} \oplus \mathcal{V} \oplus \mathcal{V} \vee \mathcal{V} \oplus \dots \oplus \mathcal{V} \vee \mathcal{V} \vee \dots \vee \mathcal{V} + \dots, \\ &= \bigoplus_{n=0}^{\infty} (\vee^n \mathcal{V}); \\ \mathcal{V} &:= \sum_{\mathbf{k} \in \pi \mathbb{Z}_2^L} \oplus V_{\mathbf{k}}. \end{aligned} \quad (3.22)$$

The linear operators which can be used to construct the branching operator are defined in terms of the so-called creation and annihilation operators on  $\mathcal{F}^+$ . For  $v_{\mathbf{k}} \in V_{\mathbf{k}}$ ,  $v^{\mathbf{k}*} \in V_{\mathbf{k}}^*$  define the operators  $a^\dagger(v_{\mathbf{k}}) : \vee^n \mathcal{V} \rightarrow \vee^{n+1} \mathcal{V}$ ,  $a(v^{\mathbf{k}*}) : \vee^n \mathcal{V} \rightarrow \vee^{n-1} \mathcal{V}$  by (see, for example, [22])

$$a^\dagger(v_{\mathbf{k}}) \cdot v_{\mathbf{k}_1} \vee v_{\mathbf{k}_2} \vee \dots \vee v_{\mathbf{k}_n} = v_{\mathbf{k}} \vee v_{\mathbf{k}_1} \vee v_{\mathbf{k}_2} \vee \dots \vee v_{\mathbf{k}_n};$$

<sup>4</sup> A similar situation may apply in the antisymmetric case, but we shall not consider it further here.

$$a(v^{\mathbf{k}^*}) \cdot v_{\mathbf{k}_1} \vee v_{\mathbf{k}_2} \vee \cdots \vee v_{\mathbf{k}_n} = \sum_{m=1}^n \delta_{\mathbf{k}_m}^{\mathbf{k}^*} v^{\mathbf{k}^*}(v_{\mathbf{k}_m}) v_{\mathbf{k}_1} \vee \cdots \widehat{v}_{\mathbf{k}_m} \cdots \vee v_{\mathbf{k}_n}$$

where  $\widehat{v}$  denotes the omission of the corresponding vector (the dual action has been formally extended to be zero on differently labelled spaces, and the corresponding  $\delta_{\mathbf{k}_m}^{\mathbf{k}^*}$  factor displayed explicitly). The operators so defined should then be formally summed to give operators on the whole of  $\mathcal{F}^+$  (and by definition  $a^{\mathbf{k}}(v^*) \cdot \mathbb{C} = 0$ ), for which we retain the same symbol. In particular for the unit vectors  $e_{\mathbf{k}\alpha}$  and their duals  $e^{\mathbf{k}\alpha}$  we define

$$a^\dagger(e_{\mathbf{k}\alpha}) := a_{\mathbf{k}\alpha}^\dagger, \quad a(e^{\mathbf{k}\alpha}) := a^{\mathbf{k}\alpha}. \tag{3.23}$$

The operators  $a(u^*), a^\dagger(v)$  fulfil the commutation (ordering) relations  $a(u^*)a^\dagger(v) - a^\dagger(v)a(u^*) \equiv [a(u^*), a^\dagger(v)] = u^*(v)\mathbb{1}$ , where  $\mathbb{1}$  is the unit operator on  $\mathcal{F}^+$ ; in particular for the mode operators  $a_{\mathbf{k}\alpha}^\dagger, a^{\mathbf{k}\alpha}$  we have

$$[a^{\mathbf{k}\alpha}, a_{\mathbf{l}\beta}^\dagger] = \delta_{\mathbf{l}}^{\mathbf{k}} \delta_{\beta}^{\alpha} \mathbb{1}. \tag{3.24}$$

Moreover, if we define the ‘ground’ state to be  $\mathbb{1} \in \mathbb{C}$ , we have the algebraic means to write an *arbitrary* element of the corresponding distinguished basis in Fock space,

$$e_{\mathbf{k}_1\alpha_1} \vee e_{\mathbf{k}_2\alpha_2} \vee \cdots \vee e_{\mathbf{k}_n\alpha_n} := a_{\mathbf{k}_1\alpha_1}^\dagger \cdot a_{\mathbf{k}_2\alpha_2}^\dagger \cdots a_{\mathbf{k}_n\alpha_n}^\dagger \cdot \mathbb{1}. \tag{3.25}$$

In what follows, it will be notationally more compact to introduce the so-called Dirac bra–ket notation for vectors in  $\mathcal{V}$  and their duals. Thus formally we write

$$\begin{aligned} \mathbb{1} &\leftrightarrow |0\rangle, & \mathbb{1}^* &\leftrightarrow \langle 0|; \\ e_{\mathbf{k}\alpha} &\leftrightarrow a_{\mathbf{k}\alpha}^\dagger |0\rangle \equiv |\mathbf{k}, \alpha\rangle, & e^{*\mathbf{k}\alpha} &\leftrightarrow \langle 0| a^{\mathbf{k}\alpha} \equiv \langle \mathbf{k}, \alpha|; \\ e_{\mathbf{k}_1\alpha_1} \vee e_{\mathbf{k}_2\alpha_2} \vee \cdots \vee e_{\mathbf{k}_n\alpha_n} &:= a_{\mathbf{k}_1\alpha_1}^\dagger \cdot a_{\mathbf{k}_2\alpha_2}^\dagger \cdots a_{\mathbf{k}_n\alpha_n}^\dagger |0\rangle \equiv |\mathbf{k}_1\alpha_1, \mathbf{k}_2\alpha_2, \dots, \mathbf{k}_n\alpha_n\rangle, \end{aligned}$$

where the latter list may include repetition. In this case, the explicit notation

$$|\mathbf{k}_1\alpha_1, m_1; \mathbf{k}_2\alpha_2, m_2; \dots \mathbf{k}_r\alpha_r, m_r\rangle = (a_{\mathbf{k}_1\alpha_1}^\dagger)^{m_1} \cdot (a_{\mathbf{k}_2\alpha_2}^\dagger)^{m_2} \cdots (a_{\mathbf{k}_r\alpha_r}^\dagger)^{m_r} |0\rangle, \tag{3.26}$$

(corresponding to the so-called *number basis*) is occasionally mandatory. Finally, we introduce the natural Cartesian inner product on these state vectors (with the  $e_{\mathbf{k}\alpha}$  orthonormal in  $\mathcal{V}$ ), extended to  $\mathcal{F}^+$  in such a way that each creation and annihilation pair is mutually Hermitian, and in general

$$\begin{aligned} &\langle \mathbf{k}_1\alpha_1, m_1; \mathbf{k}_2\alpha_2, m_2; \dots \mathbf{k}_r\alpha_r, m_r | \mathbf{l}_1\beta_1, n_1; \mathbf{l}_2\beta_2, n_2; \dots \mathbf{l}_s\beta_s, n_s \rangle \\ &= \delta_{rs} \prod_{q=1}^r \delta_{\mathbf{k}_q\mathbf{l}_q} \delta_{\alpha_q\beta_q} \cdot \delta_{m_q n_q} m_q! \end{aligned} \tag{3.27}$$

Although the general structure will be needed in the formalism below, sample states are in practice those belonging to a fixed number  $n$  of subsystems (for example  $n = L$ , the number of taxa), with (distinct) labelled momenta without multiplicity, of the general form

$$|P\rangle = \sum_{\alpha_1, \alpha_2, \dots, \alpha_n=0}^{K-1} P^{\alpha_1\alpha_2\cdots\alpha_n} |\mathbf{k}_1\alpha_1, \mathbf{k}_2\alpha_2, \dots, \mathbf{k}_n\alpha_n\rangle. \tag{3.28}$$

Such state vectors can immediately be attributed to a theoretical probability density for  $n$  taxa provided that the coefficients are positive and that their sum is unity. For technical reasons, we introduce an auxiliary ‘reservoir’ state (dual or ‘bra’ vector)

$${}^{(n)}\langle \Omega| = \sum_{\alpha_1, \alpha_2, \dots, \alpha_n=0}^{K-1} \langle \mathbf{k}_1\alpha_1, \mathbf{k}_2\alpha_2, \dots, \mathbf{k}_n\alpha_n|$$

so that this condition can be written as

$${}^{(n)}\langle \Omega | P \rangle = 1 \iff \sum_{\alpha_1, \alpha_2, \dots, \alpha_n=0}^{K-1} P^{\alpha_1 \alpha_2 \dots \alpha_n} = 1. \tag{3.29}$$

In full generality, the auxiliary vector (allowing for multiplicities and summing over different momenta) becomes

$$\langle \Omega | := \langle 0 | \exp \left( \sum_{\mathbf{k} \in \pi \mathbb{Z}_2^L} \sum_{\alpha=0}^{K-1} a^{\mathbf{k}\alpha} \right), \tag{3.30}$$

or

$${}^\chi \langle \Omega | = \langle 0 | \exp \left( \sum_{\mathbf{k} \in \pi \mathbb{Z}_2^L} \sum_{\alpha=0}^{K-1} \chi_{\mathbf{k}\alpha} a^{\mathbf{k}\alpha} \right), \tag{3.31}$$

where the latter form is convenient for notational purposes (with the understanding that  $\chi_{\mathbf{k}\alpha} \rightarrow 1$ ).

We shall be concerned with general functions  $f$  (such as the probabilities  $P$ , and below with operators built from the creation and annihilation mode operators) which are obtained as formal sums of terms depending on the ‘momentum’ labels, say  $f_{\mathbf{k}l\dots}$ . With the convention we have adopted (of scaling the  $\mathbf{k}$  by  $\pi$ ) with any such function we can associate functions over a dual space  $\mathbf{x}, \mathbf{y} \dots \in \mathbb{Z}_2^L$  by a formal Fourier transform. This is of course the discrete Fourier–Hadamard transformation (the phase factors are simply  $\pm 1$ ), and the functions  $f$  on ‘configuration’ (position) space are periodic with periods  $2\mathbf{a}$  for  $\mathbf{a} \in \mathbb{Z}_2^L$ . In particular for the constant function in one variable  $\mathbb{1}_{\mathbf{k}} = 1$ ,

$$\delta(\mathbf{x}) = \left( \frac{1}{2^L} \right) \sum_{\mathbf{k} \in \pi \mathbb{Z}_2^L} e^{i(\mathbf{k}\cdot\mathbf{x})}, \quad 1 = \sum_{\mathbf{x} \in \mathbb{Z}_2^L} \delta(\mathbf{x}) e^{-i(\mathbf{k}\cdot\mathbf{x})},$$

(where  $\delta(\mathbf{x}) = \delta(\mathbf{x}, \mathbf{0})$ ). More generally,

$$f(\mathbf{x}) := \left( \frac{1}{2^L} \right) \sum_{\mathbf{k} \in \pi \mathbb{Z}_2^L} f_{\mathbf{k}} e^{i(\mathbf{k}\cdot\mathbf{x})}, \quad f_{\mathbf{k}} = \sum_{\mathbf{x} \in \mathbb{Z}_2^L} f(\mathbf{x}) e^{-i(\mathbf{k}\cdot\mathbf{x})}. \tag{3.32}$$

In two variables, we have in turn

$$f(\mathbf{x}, \mathbf{y}) := \left( \frac{1}{2^L} \right)^2 \sum_{\mathbf{k} \in \pi \mathbb{Z}_2^L} \sum_{\mathbf{l} \in \pi \mathbb{Z}_2^L} f_{\mathbf{k}l} e^{i(\mathbf{k}\cdot\mathbf{x} + \mathbf{l}\cdot\mathbf{y})}, \quad f_{\mathbf{k}l} = \sum_{\mathbf{x} \in \mathbb{Z}_2^L} f(\mathbf{x}, \mathbf{y}) e^{-i(\mathbf{k}\cdot\mathbf{x} + \mathbf{l}\cdot\mathbf{y})}, \tag{3.33}$$

and generally

$$f(\mathbf{x} + 2\mathbf{a}, \mathbf{y} + 2\mathbf{b}, \dots) = f(\mathbf{x}, \mathbf{y}, \dots).$$

As an example of the creation and annihilator formalism, let us give an operator on  $\mathcal{F}^+$  equivalent to the branching operator  $\delta : V \rightarrow V \otimes V$  introduced above (which has to be extended case by case to allow for branchings on particular factors of  $\otimes^n V$  for a particular tree). Recall that the general form

$$\delta(e_\alpha) = \Gamma_\alpha^{\beta\gamma} e_\beta \otimes e_\gamma \tag{3.34}$$

was subsequently specialized to  $\Gamma_\alpha^{\beta\gamma} = \delta_\beta^\alpha \delta_\gamma^\alpha$  on the basis of conditional independence. Next assume that the copies of  $V$  involved are distinguished by different labels  $\mathbf{k}, \mathbf{l}, \mathbf{m}$  so that there is no difference between the above use of  $\otimes$  and the correct  $\vee$  as far as the symmetric algebra is concerned (below we shall see that the momentum labels are such that  $\mathbf{k} = \mathbf{l} + \mathbf{m}$ ). Consider then the operator  $\Delta = \sum_{\alpha, \beta, \gamma} \Gamma_\alpha^{\beta\gamma} a^\alpha a^\dagger_\beta a^\dagger_\gamma$ , and its action on a state  $V \ni |p\rangle = p^0|0\rangle + p^1|1\rangle + \dots + p^{K-1}|K-1\rangle = \sum_\xi p^\xi |\xi\rangle$ :

$$\begin{aligned}
 \Delta|p\rangle &= \sum_{\alpha,\beta,\gamma} \Gamma_{\alpha}^{\beta\gamma} a^{\dagger}_{\beta} a^{\dagger}_{\gamma} a^{\alpha} \sum_{\xi} p^{\xi} a^{\dagger}_{\xi} |0\rangle \\
 &= \sum_{\alpha,\beta,\gamma} \Gamma_{\alpha}^{\beta\gamma} a^{\dagger}_{\beta} a^{\dagger}_{\gamma} \sum_{\xi} p^{\xi} [a^{\alpha}, a^{\dagger}_{\xi}] |0\rangle \\
 &= \sum_{\alpha,\beta,\gamma} \Gamma_{\alpha}^{\beta\gamma} a^{\dagger}_{\beta} a^{\dagger}_{\gamma} p^{\alpha} |0\rangle \\
 &= \sum_{\alpha,\beta,\gamma} [p^{\alpha} \Gamma_{\alpha}^{\beta\gamma}] |\beta, \gamma\rangle.
 \end{aligned}
 \tag{3.35}$$

Thus, indeed, the requisite branching from the initial ancestral density  $|p\rangle = \sum_{\alpha} p^{\alpha} |\alpha\rangle$  to the density for two taxa after branching, with characters shared equally ( $|P\rangle = \sum_{\alpha} p^{\alpha} |\alpha, \alpha\rangle$  for the special choice (2.16) of  $\Gamma$ , has been effected, and the operator  $\Delta$  provides a generalization of the splitting operator  $\delta$  of (2.10), (2.17) suitable for representing embeddings of the latter on individual factors of the tensor product, as in (2.18). In section 6 below, we return to this operator in the context of a dynamical change model for branching. As will be seen, it needs to be embellished by edge ‘momentum’ labels in order to generate appropriate phylogenetic trees, and also to be assigned a time dependence corresponding to the fact that branching events will occur at specific times in an evolutionary sense. These apparent complications need to be contrasted with the fact that if the splitting operator  $\delta$  is used in its original form, for a specific tree, its action on tensor products must be extended on a case-by-case basis, as in (2.18). In sections 4, 5 below, we turn to a review of the path integral method for solving the time evolution of systems described in the operator language, and then apply the technique to a system of taxa which is ‘free’, that is, evolving without any phylogenetic association, after having developed the appropriate form for the rate operator of such a system.

#### 4. Review of path integral formalism

In this section, we review briefly the path integral formalism for the representation of the time development of stochastic processes whose ‘microscopic’ states represent probabilities of certain ‘particle’ numbers at each time. The aim of the next section will be to apply the technique to the multilinear representation of taxonomic states developed in section 2 and transcribed into the ‘occupation number’ representation in section 3 above. The task at hand is to transcribe the abstract occupation number representation (as developed for our purposes in the previous section) into a formalism of integral operators acting on generating functions representing the appropriate probability densities. This section closely follows the presentation of Peliti [18].

For a single system we therefore have microscopic states of the form (see (3.23))  $|n\rangle = a^{\dagger n} |0\rangle$ , with the creation and annihilation operators  $a^{\dagger}$  and  $a$  being Hermitian conjugates of each other with

$$\begin{aligned}
 a^{\dagger}|n\rangle &= |n + 1\rangle, & a|n\rangle &= |n - 1\rangle, \\
 \langle n|m\rangle &= n! \delta_{mn}.
 \end{aligned}
 \tag{4.36}$$

Next we make the transcription from states

$$|\phi\rangle = \sum_{n=0}^{\infty} \phi^n |n\rangle, \quad \langle\phi|\psi\rangle = \sum_{n=0}^{\infty} n! \phi_n \psi_n$$

to a space of functions

$$|\phi\rangle \leftrightarrow \phi(z) = \sum_{n=0}^{\infty} \phi_n z^n,$$

where the variable ‘ $z$ ’ is a formal variable if (as in the usual statistical context)  $\phi(z)$  is meant as a formal generating function. However, for the present development it is convenient to allow  $z$  to be complex and to regard the  $\phi(z)$  as analytic functions belonging to a Hilbert space. In terms of defining path integrals  $z$  can be taken to be real, or analytically continued subject to certain prescribed asymptotic behaviour (possibly together with constraints forcing its passage through specified points of the complex plane).

Using the elementary identity

$$n! \delta_{mn} = \int dz z^n \left(-\frac{d}{dz}\right)^m \delta(z) \tag{4.37}$$

(which can be established by integration by parts) the scalar product (4.36) becomes

$$\langle \phi | \psi \rangle = \int dz \phi(z) \psi \left(-\frac{d}{dz}\right) \delta(z), \quad \text{or} \quad \langle \phi | \psi \rangle = \int \frac{dz d\zeta}{2\pi} \phi(z) \psi(i\zeta) e^{-iz\zeta}. \tag{4.38}$$

Associated with the matrix elements  $\mathcal{A}_{mn} = \langle m | A | n \rangle$  of any operator in the number basis is the integral kernel  $\mathcal{A}(\bar{z}, \zeta)$

$$\mathcal{A}(z, \zeta) = \sum_{m,n=0}^{\infty} \frac{z^m}{m!} A_{mn} \frac{\zeta^n}{n!} \tag{4.39}$$

such that

$$|\psi\rangle = A|\phi\rangle = \sum_{m,n} \frac{|m\rangle\langle m|}{m!} A \frac{|n\rangle\langle n|}{n!} |\phi\rangle$$

can be expressed via  $\psi(z) = \sum \psi_m z^m$ , with

$$\psi(z) = \int \frac{d\zeta d\zeta'}{2\pi} \mathcal{A}(z, \zeta) \phi(i\zeta') e^{-i\zeta\zeta'} \tag{4.40}$$

easily established using the identity (4.37) above. Similarly, the integral kernel of the product  $\mathcal{A}\mathcal{B}$  of two operators  $\mathcal{A}, \mathcal{B}$  is

$$\mathcal{A}\mathcal{B}(z, \zeta) = \int \frac{d\eta d\eta'}{2\pi} \mathcal{A}(z, \eta) \mathcal{B}(i\eta', \zeta) e^{-i\eta\eta'}. \tag{4.41}$$

The integral kernel  $\mathcal{A}(z, \zeta)$  has a natural combinatorial connection to the normal kernel  $A(z, \zeta)$  where  $A$  is expressed in terms of creation and annihilation operators:

$$A = \sum_{m,n=0}^{\infty} a^{\dagger m} A_{mn} a^n, \quad \text{define} \quad A(z, \zeta) := \sum_{m,n} z^m A_{mn} \zeta^n.$$

Then there is the simple relationship

$$\mathcal{A}(z, \zeta) = e^{z\zeta} A(z, \zeta).$$

Consider the effect of stochastic time evolution on the system. In the linear case, the state probabilities are assumed to change according to the master equation

$$\frac{d}{dt} \phi_n = \sum_{m \neq n} (r_{m \rightarrow n} \phi_m - r_{n \rightarrow m} \phi_n) \tag{4.42}$$

where  $r_{m \rightarrow n}$  are the transition rates. It is convenient to define  $R_{nm} = r_{m \rightarrow n}$  and  $R_{nn} = -\sum_m r_{n \rightarrow m} = -\sum_m R_{mn}$  so that the time evolution becomes

$$\frac{d}{dt}\phi_n = \sum_m R_{nm}\phi_m \tag{4.43}$$

with the understanding that the rate matrix satisfies  $\sum_n R_{nm} = 0$ , for all  $m$ , or introducing the reservoir state  $\langle \Omega |$  from above, and regarding  $R(t)$  as an operator on state space which can be time dependent<sup>5</sup>,

$$\frac{d}{dt}|\phi(t)\rangle = R(t)|\phi(t)\rangle, \quad \text{with} \quad \langle \Omega | R(t) = 0. \tag{4.44}$$

With the above notation we can now develop a path integral representation for the evolution kernel of the system. Approximate the form of the evolution operator for a small change as  $M_{(t+\delta t, t)} \simeq e^{R(t)\delta t}$ , and for the evolution operator as a whole as a product of infinitesimal changes

$$M_{(T,0)} \simeq M_{(T, T-\delta t)} \cdot M_{(T-\delta t, T-2\delta t)} \cdots M_{(2\delta t, \delta t)} \cdot M_{(\delta t, 0)}.$$

Approximating each of the exponentials by a linear expression, and using the above relation between normal and integral kernels leads to

$$\mathcal{M}_{(t+\delta t, t)}(z, \zeta) \simeq e^{z\zeta} (1 + \delta t R(t)(z, \zeta)).$$

Using this and iterating (4.41) to give a multiple integral representation of this product, assuming  $T = N\delta t$ , we have

$$\begin{aligned} M_{(T,0)}(z, \zeta) &\simeq \int M_{(T, T-\delta t)}(z, \eta_1) \frac{d\eta_1 d\eta'_1 e^{-i\eta_1 \eta'_1}}{2\pi} M_{(T-\delta t, T-2\delta t)}(\eta'_1, \eta_2) \frac{d\eta_2 d\eta'_2 e^{-i\eta_2 \eta'_2}}{2\pi} \dots \\ &\quad \cdot M_{(2\delta t, \delta t)}(\eta'_{N-2}, \eta_{N-1}) \frac{d\eta_{N-1} d\eta'_{N-1} e^{-i\eta_{N-1} \eta'_{N-1}}}{2\pi} M_{(\delta t, 0)}(\eta'_{N-1}, \zeta) \\ &\simeq \int \prod_{\ell=0}^{N-1} \frac{d\eta_\ell d\eta'_\ell}{2\pi} \cdot \exp\left(\sum_{\ell=0}^{N-1} [-i\eta'_{\ell+1}(\eta_{\ell+1} - \eta_\ell) + \delta t R(t)(i\eta'_{\ell+1}, \eta_\ell)]\right) \cdot e^{z\eta_N} \end{aligned} \tag{4.45}$$

which leads formally in the limit  $N \rightarrow \infty$  to the path integral representation (cf [18] equations (2.23), (2.24))

$$\mathcal{M}_T(z, \zeta) = \int d[\eta] d[\eta'] \exp\left(\int_0^T dt (-i\eta'(t)\dot{\eta}(t) + iR_t(i\eta', \eta)) + z\eta(T)\right). \tag{4.46}$$

Here the  $2\pi$  factors have been incorporated into the path integral measure, and the integrations over paths  $\eta(t), \eta'(t)$  from 0 to  $T$  are made with the boundary conditions on each endpoint given by

$$\eta(0) = \zeta, \quad i\eta'(T) = z. \tag{4.47}$$

The additional boundary term  $\exp(z\eta(T))$  also arises from the continuum ( $N \rightarrow \infty$ ) limit of the iterated product representation.

It is important to point out that the path integral representation [5, 6, 18] also allows closed form expressions to be written down for the means (and in principle higher moments) of any desired observable quantities. This has not only formal significance but also, depending on the operator, opens an avenue for explicit analytical calculations.

<sup>5</sup> This entails  $d\phi_n/dt = \langle n|R|\phi\rangle/n!$ , consistent with the resolution of the identity (see (4.38) above).

### 5. Evolution kernel for free phylogenetic system

With our review of path integrals for stochastic systems in hand, we now return to the discussion of phylogenetic systems in the notation of section 3. We concentrate here on the ‘free’ system, that is, phylogenetic evolution without phylogenetic branching. As we now argue, the normal kernel of the rate operator can be taken to be quadratic, so that the entire path integral assumes the Gaussian form and admits a formal steepest descent evaluation. In the next section, we also introduce interactions along the lines of (3.35) and indicate in simple examples which indeed reproduce the expected evolution (at least if the rate matrix is time independent) that this leads to the correct probabilities<sup>6</sup>.

Transition rates have been discussed in section 2 in the tensor formalism, and in section 4 in introducing the path integral representation. However in the context of section 3, the appropriate time evolution must be the assignment of a rate operator to each possible edge ‘momentum’ label,  $\mathbf{k} \in \pi\mathbb{Z}_2^L$ . In contrast to (4.42) then, in which it is assumed that the rates  $r_{m \rightarrow n}$  and the  $|m\rangle, |n\rangle$  refer to *differing* occupation numbers, we thus construct initially a *number-conserving* rate operator, at least inasmuch as the ‘particle number’ operator does not distinguish between the character types  $\alpha = 0, 1, \dots, K - 1$  which take on the status of ‘internal’ degrees of freedom. Indeed the number operator for edge  $\mathbf{k}$  is

$$\begin{aligned}
 N_{\mathbf{k}} &= \sum_{\alpha} a_{\mathbf{k}\alpha}^{\dagger} a^{\mathbf{k}\alpha}, \quad \text{such that} \\
 [N_{\mathbf{k}}, a_{\mathbf{k}\alpha}^{\dagger}] &= a_{\mathbf{k}\alpha}^{\dagger}, \quad [N_{\mathbf{k}}, a^{\mathbf{k}\alpha}] = -a^{\mathbf{k}\alpha}, \\
 [N_{\mathbf{k}}, a_{\mathbf{k}\alpha}^{\dagger} a^{\mathbf{k}\beta}] &= 0 \quad \forall \alpha, \beta
 \end{aligned}
 \tag{5.48}$$

This means of course that the rate operator must be bilinear in both creation and annihilation operators of type  $\mathbf{k}$ , leading to the second-quantized expression

$$R = \sum_{\mathbf{k} \in \pi\mathbb{Z}_2^L} R_{\mathbf{k}} = \sum_{\mathbf{k} \in \pi\mathbb{Z}_2^L} \sum_{\alpha, \beta} a_{\mathbf{k}\alpha}^{\dagger} R_{\mathbf{k}}^{\alpha \beta} a^{\mathbf{k}\beta}.
 \tag{5.49}$$

As mentioned in section 3 above, we will be concerned with *single* occupation numbers for each momentum mode (for generalizations, see the concluding remarks in section 7 below). Thus for a general tensor state (3.28) (cf section 2),

$$|\dot{P}\rangle = R|P\rangle = \sum_{\gamma_1, \gamma_2, \dots, \gamma_n=0}^{K-1} P^{\gamma_1 \gamma_2 \dots \gamma_n} R|\mathbf{k}_1 \gamma_1, \mathbf{k}_2 \gamma_2, \dots, \mathbf{k}_n \gamma_n\rangle.$$

Using the fundamental relation

$$\begin{aligned}
 [R, a_{\mathbf{l}\gamma}^{\dagger}] &= \left[ \sum_{\mathbf{k} \in \pi\mathbb{Z}_2^L} \sum_{\alpha, \beta} a_{\mathbf{k}\alpha}^{\dagger} R_{\mathbf{k}}^{\alpha \beta} a^{\mathbf{k}\beta}, a_{\mathbf{l}\gamma}^{\dagger} \right] \\
 &= \sum_{\mathbf{k} \in \pi\mathbb{Z}_2^L} \sum_{\alpha, \beta} a_{\mathbf{k}\alpha}^{\dagger} R_{\mathbf{k}}^{\alpha \beta} [a^{\mathbf{k}\beta}, a_{\mathbf{l}\gamma}^{\dagger}] \\
 &= \sum_{\alpha} a_{\mathbf{l}\alpha}^{\dagger} R_{\mathbf{l}}^{\alpha \gamma},
 \end{aligned}
 \tag{5.50}$$

<sup>6</sup> The formalism also applies to the inhomogeneous case (time-dependent rates), provided that the ‘propagator’ is known (see below).

where (3.24) has been used, together with (3.26), we find finally as required

$$|\dot{P}\rangle \equiv \sum_{\gamma_1, \gamma_2, \dots, \gamma_n=0}^{K-1} \dot{P}^{\gamma_1 \gamma_2 \dots \gamma_n} |\mathbf{k}_1 \gamma_1, \mathbf{k}_2 \gamma_2, \dots, \mathbf{k}_n \gamma_n\rangle, \quad \text{where} \tag{5.51}$$

$$\dot{P}^{\gamma_1 \gamma_2 \dots \gamma_n} = \sum_{\gamma} \left( R_{\mathbf{k}_1}^{\gamma_1} P^{\gamma \gamma_2 \dots \gamma_n} + R_{\mathbf{k}_2}^{\gamma_2} P^{\gamma_1 \gamma \dots \gamma_n} + \dots + R_{\mathbf{k}_n}^{\gamma_n} P^{\gamma_1 \gamma_2 \dots \gamma} \right),$$

whence

$$P^{\gamma_1 \gamma_2 \dots \gamma_n}(T) = \sum_{\delta_i} M_T^{\gamma_1}_{\delta_1} M_T^{\gamma_2}_{\delta_2} \dots M_T^{\gamma_n}_{\delta_n} P^{\delta_1 \delta_2 \dots \delta_n}, \quad \text{where } M_T^{\gamma_i}_{\delta_i} = (e^{T R_{\mathbf{k}_i}})^{\gamma_i}_{\delta_i}. \tag{5.52}$$

It remains to transcribe these results into the path integral notation (4.46) and verify that the same time evolution is predicted in the ‘free’ (Gaussian) case at least for time-independent rates. Clearly, for each degree of freedom  $\mathbf{k}, \alpha$  there is a pair of classical paths  $\eta'(t)_{\mathbf{k}\alpha}, \eta(t)^{\mathbf{k}\beta}$  or collectively simply  $\eta'(t), \eta(t)$ . From the fact that the rate operator is expressed by (5.49) in the normal form, we take

$$R_t(i\eta', \eta) = \sum_{\mathbf{k}, \alpha, \beta} i\eta'(t)_{\mathbf{k}\alpha} R_{\mathbf{k}}^{\alpha} \eta(t)^{\mathbf{k}\beta}.$$

In these circumstances the time evolution kernel is particularly simple. Explicitly,

$$\mathcal{M}_T(z', \zeta) = \int [d\eta][d\eta'] \exp \left( \int_0^T dt \left( -i \sum_{\mathbf{k}, \alpha} \eta'^{\mathbf{k}\alpha}(t) \dot{\eta}_{\mathbf{k}\alpha}(t) + i \sum_{\mathbf{k}, \alpha, \beta} \eta'(t)_{\mathbf{k}\alpha} R_{\mathbf{k}}^{\alpha} \eta(t)^{\mathbf{k}\beta} \right) + \sum_{\mathbf{k}, \alpha} z'_{\mathbf{k}\alpha} \eta(T)^{\mathbf{k}\alpha} \right), \tag{5.53}$$

subject to the appropriate boundary conditions. The integration over all paths  $\eta'(t)$ , imposes a functional- $\delta$  constraint on  $\eta(t)$ , namely

$$\dot{\eta}^{\mathbf{k}\alpha}(t) = \sum_{\mathbf{k}, \beta} R_{\mathbf{k}}^{\alpha} \eta(t)^{\mathbf{k}\beta}, \quad \text{whence } \eta^{\mathbf{k}\alpha}(T) = \sum_{\mathbf{k}, \beta} M_T^{\alpha}_{\mathbf{k}\beta} \eta(0)^{\mathbf{k}\beta} \tag{5.54}$$

(if the transition rates are time independent), where  $M_{T\mathbf{k}}$  is given by (5.51) above. Thus the  $\eta(t)$  path integral contribution (up to a normalization constant<sup>7</sup>) comes from the boundary term which gives using (4.47)

$$\mathcal{M}_T(z', \zeta) = \mathcal{C} \exp \left( \sum_{\mathbf{k}\alpha} z'_{\mathbf{k}\alpha} M_T^{\alpha}_{\mathbf{k}\beta} \zeta^{\mathbf{k}\beta} \right) \tag{5.55}$$

for some integration constant  $\mathcal{C}$ .

For a phylogenetic tensor state (as discussed above, with the unit occupation number in momentum modes  $\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_n$ ) we set

$$P_t(z'_1, z'_2, \dots, z'_n) = \sum_{\alpha_1, \alpha_2, \dots, \alpha_n=0}^{K-1} P_t^{\alpha_1 \alpha_2 \dots \alpha_n} z'_{\mathbf{k}_1 \alpha_1} z'_{\mathbf{k}_2 \alpha_2} \dots z'_{\mathbf{k}_n \alpha_n}. \tag{5.56}$$

<sup>7</sup> For this case, the discrete version can be worked out explicitly as a (multi-dimensional) standard Gaussian integral, and the limit  $N \rightarrow \infty$  considered. In the present heuristic discussion, we simply assume that the steepest descent method yields the correct result.



Denoting the state collectively as  $P_t(z')$  we then have from (4.40)

$$P_T(z') = \int \prod_{\mathbf{k}, \alpha} [d\zeta'_{\mathbf{k}\alpha}] [d\zeta^{\mathbf{k}\alpha}] \mathcal{M}_T(z', \zeta) e^{-i \sum_{\mathbf{k}, \alpha} \zeta'_{\mathbf{k}\alpha} \zeta^{\mathbf{k}\alpha}} P_0(i\zeta'). \tag{5.57}$$

Substituting (5.55) it is evident that the  $\zeta$  integrations impose a functional- $\delta$  constraint<sup>8</sup>  $\delta(i\zeta' - z' \cdot M_T)$ , or

$$\begin{aligned} P_T(z') &= P_0(z' \cdot M_T), \quad \text{or} \\ P_T(z'_1, z'_2, \dots, z'_n) &= \sum_{\alpha_1, \alpha_2, \dots, \alpha_n=0}^{K-1} P_0^{\alpha_1 \alpha_2 \dots \alpha_n} z'_1 \cdot M_{T \mathbf{k}_1 \mathbf{k}_1 \alpha_1} z'_2 \cdot M_{T \mathbf{k}_2 \mathbf{k}_2 \alpha_2} \dots z'_n \cdot M_{T \mathbf{k}_n \mathbf{k}_n \alpha_n}, \quad \text{or} \\ P_T(z'_1, z'_2, \dots, z'_n) &= \sum_{\alpha_1, \alpha_2, \dots, \alpha_n=0}^{K-1} P_T^{\alpha_1 \alpha_2 \dots \alpha_n} z'_{\mathbf{k}_1 \alpha_1} z'_{\mathbf{k}_2 \alpha_2} \dots z'_{\mathbf{k}_n \alpha_n}, \quad \text{where finally} \\ P_T^{\gamma_1 \gamma_2 \dots \gamma_n} &= \sum_{\delta_i} M_T^{\gamma_1}_{\delta_1} M_T^{\gamma_2}_{\delta_2} \dots M_T^{\gamma_n}_{\delta_n} P_0^{\delta_1 \delta_2 \dots \delta_n} \end{aligned} \tag{5.58}$$

as derived explicitly above.

It is instructive to rewrite the classical ‘free’ evolution kernel in Fourier transform space. Recalling the duality between ‘momentum’ labels  $\mathbf{k}, \mathbf{l} \in \pi \mathbb{Z}_2^L$  and ‘position’ coordinates  $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_2^L$ , define

$$\begin{aligned} \eta(t)^{\mathbf{k}\alpha} &= \sum_{\mathbf{x}} \eta^\alpha(\mathbf{x}, t) e^{-i\mathbf{k}\cdot\mathbf{x}}, & \eta'(t)_{\mathbf{k}\alpha} &= \frac{1}{2^L} \sum_{\mathbf{y}} \eta'_\alpha(\mathbf{y}, t) e^{i\mathbf{k}\cdot\mathbf{y}}, \\ R(t)_{\mathbf{k}\beta}^\alpha &= \sum_{\mathbf{z}} R(\mathbf{z}, t)^\alpha_\beta e^{i\mathbf{k}\cdot\mathbf{z}}, & \text{so that (5.53) becomes} \end{aligned}$$

$$\begin{aligned} \mathcal{M}_T(z', \zeta) &= \int [d\eta][d\eta'] \exp \left( \int_0^T dt \left( -i \sum_{\mathbf{x}} \eta'_\alpha(\mathbf{x}, t) \dot{\eta}^\alpha(\mathbf{x}, t) \right. \right. \\ &\quad \left. \left. + i \sum_{\mathbf{x}, \mathbf{y}} \eta'_\alpha(\mathbf{x}, t) R(\mathbf{x} - \mathbf{y}, t)^\alpha_\beta \eta^\alpha(\mathbf{x}, t) \right) + \sum_{\mathbf{x}} z'_\alpha(\mathbf{x}) \eta^\alpha(\mathbf{x}, T) \right). \end{aligned} \tag{5.59}$$

### 6. Interaction terms and simple examples

In this section, we turn to the complete phylogenetic system incorporating ‘interaction’ terms. In the previous section, we constructed the ‘free’ part of the evolution kernel  $\mathcal{M}_T = \int [d\eta][d\eta'] \exp S_0[\eta, \eta']$  for the phylogenetic ‘fields’  $\eta'_\alpha(\mathbf{x}, t), \eta^\alpha(\mathbf{x}, t)$ . Incorporating interactions, the kernel will acquire additional trilinear terms  $S_1$  in the exponent representing phylogenetic branching events, in such a way that the manifest translation symmetry in ‘position’ space is preserved.

In section 3 above, it was pointed out that the ‘branching operator’  $\delta$  which was formally introduced in (2.17) of section 2 can be represented by a trilinear  $2 \leftarrow 1$  type operator in Fock space (compare (3.35) above). In the case where there are up to  $L$  extant taxonomic units labelled by binary  $L$ -vectors (edge ‘momenta’) to allow for the development of a particular ancestral binary tree, this vertex interaction must be given definite momentum labels.

<sup>8</sup> Setting the arbitrary integration constant to 1.

The labelling is of course always such that an edge  $\mathbf{k} \in \mathbb{Z}_2^L$  bifurcates to edges  $\mathbf{l}, \mathbf{m}$  with  $\mathbf{k} = \mathbf{l} + \mathbf{m}$ . Moreover, if the branching is pictured as a dynamical process, the interaction must be time dependent. The simplest possibility is that the system for  $L$  taxa will evolve as a result of a fixed number  $M$  of branchings at times  $t_I, I = 1, \dots, M$  between times  $t_0 \equiv 0$  (from which time some assumed ancestor(s) evolved) and  $t_{M+1} \equiv T$  (the final time of measurement). A means of forcing these events is via  $\delta(t - t_I)$  interactions at times  $t_1 < \dots < t_I < \dots < t_M$  (with  $0 \equiv t_0 < t_1$  and  $t_M < t_{M+1} \equiv T$ ).

With the above motivations we propose the following ‘interaction’ term  $S_1$  for the full evolution kernel  $\mathcal{M}_T = \int [d\eta][d\eta'] \exp S[\eta, \eta']$  of the phylogenetic system, where  $S = S_0 + S_1$  with  $S_0$  given by (5.53), (5.59) and

$$\begin{aligned}
 S_1 &= - \int_0^T dt \frac{1}{2} \sum_I \sum_{\mathbf{k}, \mathbf{l}, \mathbf{m}} \delta(t - t_I) \delta(\mathbf{k} - \mathbf{l} - \mathbf{m}) \eta'_\alpha(t) \eta'_\beta(t) \Gamma^{\alpha\beta}_\gamma \eta(t)^{\mathbf{k}\gamma} \\
 &= - \int_0^T dt \frac{1}{2} \sum_I \sum_{\mathbf{x}} \delta(t - t_I) \eta'_\alpha(\mathbf{x}, t) \eta'_\beta(\mathbf{x}, t) \Gamma^{\alpha\beta}_\gamma \eta(\mathbf{x}, t)^\gamma. \tag{6.60}
 \end{aligned}$$

As expected, the binary edge labelling is reflected in the manifest translation symmetry of this expression. With the complete model  $S_0 + S_1$ , the path integral formalism can now be used to construct (in a perturbation expansion, see below) the evolution kernel for the full system, and hence transition probabilities for evolution, from any initial state to any final state. In the case of phylogenetic inference, one is of course interested in evolution from an initial root edge (at time  $t = 0$ ) to (at time  $t = T$ ) an observed joint probability density for character types of  $L$  taxonomic units.

The model (5.53), (5.59), (6.60) is *generic* in the sense that an *arbitrary* (but fixed) number of branching events  $M$ , and *any* compatible branching processes for binary edges, are encoded. For connected binary trees, with a single root and  $L$  leaves one should of course admit only  $M = L - 1$   $\delta$ -function forcing terms, and adopt standard momentum labelling, for example, for  $L \leftarrow 1$  the root edge may be chosen as the binary  $L$ -vector  $(1, 1, \dots, 1)$ , and the edges the binary  $L$ -vectors  $(0, 0, \dots, 1), (0, \dots, 1, 0), (0, \dots, 1, 0, 0), \dots$  (denoted below by decimal equivalents  $\bar{1}, \bar{2}, \bar{4}, \dots$ ). For formal analysis with a specific tree, it may in fact be combinatorially more powerful to consider *all* such admissible  $L \leftarrow 1$  momentum routing schemes.

For completeness, we derive in the appendix a formal perturbative expansion [18], and give explicit Feynman rules for the present model (see table 1). The evolution kernel for  $S_0 + S_1$  is rewritten by expanding  $\exp(S_1)$  in a power series, so that the essential ingredients are specific path integrals of monomials in the phylogenetic path variables, weighted by the ‘free’ part. In turn, these moments can be reduced to functional derivatives of an extended ‘free’ kernel, with respect to ancillary ‘external’ path variables coupled by additional linear terms to the path variables which are being integrated over. The extended kernel is again quadratic and can be evaluated as a Gaussian in terms of the formal inverse bilinear form or propagator with appropriate boundary conditions (see the appendix). Moreover, the  $\delta$ -function forcing terms require the derivatives with respect to the external path variables to be evaluated at the interaction times  $t_I$ . The probabilities (pattern frequencies corresponding to *all* binary  $L$  leaf trees with evolution on edges determined by the specific edge rates  $R_{\mathbf{k}}(t)$ ) so constructed are *identical* to the usual likelihood calculation via extended leaf colourations for example. In the earlier second-quantized version (see [2]), the model was constructed using the canonical (creation and annihilation operator) formalism, and the interaction term treated in time-dependent perturbation theory. We emphasize that, although well known, the result in

**Table 1.** Feynman rules for evaluating probabilities for  $L \leftarrow 1$  scattering in the phylogenetic branching model. Trees are a combination of labelled graphical elements. Each tree contributes a term to the total likelihood or pattern frequency.  $(M_{\Delta\mathbf{k}})^{\alpha_{\mathbf{k}}}_{\beta_{\mathbf{k}}}$  is the Markov transition matrix for edge  $\mathbf{k}$  and time interval (edge length)  $\Delta$ , and  $\Gamma^{\alpha}_{\beta\beta'}$  is the vertex structure coefficient ( $\equiv \delta^{\alpha}_{\beta}\delta^{\alpha}_{\beta'}$ ). See the concluding remarks for comments on the role of the group of time reparametrizations.

| Element       | Diagram | Term   |
|---------------|---------|--|
| Internal edge |         | $(M_{\Delta\mathbf{k}})^{\alpha_{\mathbf{k}}}_{\beta_{\mathbf{k}}}$                                    |
| Root          |         | $(M_{\Delta_0\mathbf{k}_0})^{\alpha_{\mathbf{k}_0}}_{\beta_{\mathbf{k}_0}} p_0^{\beta_{\mathbf{k}_0}}$ |
| Leaf          |         | $(M_{\Delta,\mathbf{k}_i})^{\alpha_{\mathbf{k}_i}}_{\beta_{\mathbf{k}_i}}$                             |
| Vertex        |         | $\Gamma^{\alpha_{\mathbf{k}}}_{\beta_l\beta_m} \delta(\mathbf{k} - \mathbf{l} - \mathbf{m})$           |

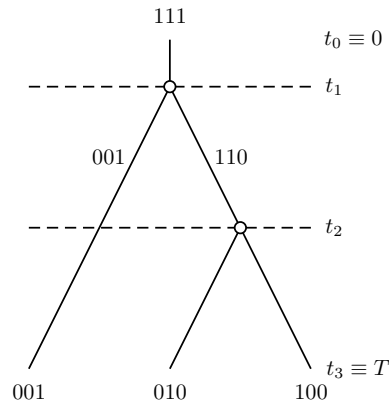
our formalism follows *automatically* from the time evolution kernel for the model (effectively, an appropriate Markov rate operator lifted to the whole Fock space), so that in this sense we have produced a truly dynamical model for phylogenetic branching processes.

We illustrate our results by reiterating some concrete examples from [2] together with some further remarks. Consider the case  $L = 3, M = 2$ . Nonzero rate constants are chosen for the root and leaf momenta  $\vec{7} = (111), \vec{1} = (001), \vec{2} = (010)$  and  $\vec{4} = (100)$  respectively, together with a *single* additional momentum  $\vec{6} = (110)$  associated with the tree  $\mathcal{T} = (\vec{1}(\vec{2}\vec{4}))$  of figure 2. Clearly, the contribution to the  $3 \leftarrow 1$  scattering probability (or likelihood) associated with this tree is, as required, the term arising (in the operator formalism [2]) from inserting intermediate states in the above with the correct intermediate edge momenta, or (in the perturbation expansion of the path integral method) from the correct linking of propagators and vertices at this order (see Feynman rules in the appendix and table 1). *Either* approach gives finally

$$\begin{aligned}
 P_{\mathcal{T}}^{\alpha_1\alpha_2\alpha_4} &= \langle \alpha_1\vec{1}, \alpha_2\vec{2}, \alpha_4\vec{4} | P_{\mathcal{T}}(T) \rangle = \langle \alpha_1\vec{1} \alpha_2\vec{2} \alpha_4\vec{4} | M_{\mathcal{T}}(T, 0) | p_{\vec{7}}(0) \rangle \\
 &= \sum M_2^{\alpha_2}_{\beta_2} M_4^{\alpha_4}_{\beta_4} \Gamma_2^{\beta_2\beta_4}_{\gamma_6} \cdot M_6^{\gamma_6}_{\beta_6} M_1^{\alpha_1}_{\beta_1} \Gamma_1^{\beta_1\beta_6}_{\beta_7} \cdot M_7^{\beta_7}_{\alpha_7} p^{\alpha_7}.
 \end{aligned}
 \tag{6.61}$$

Here  $|p_{\vec{7}}(0)\rangle = \sum p^{\alpha_7}(0)|\alpha_7\vec{7}\rangle$  is the state representing the initial root edge probability density, and the  $M_{\mathbf{k}}$  are the Markov transition matrices for the appropriate edges, namely  $M_{\mathbf{k}} = e^{\Delta_{\mathbf{k}}R_{\mathbf{k}}}$  with  $\Delta_{\mathbf{k}}$  the time evolution on edge  $\mathbf{k}$ ,  $\Delta_{\mathbf{k}} = t_{I'} - t_I$  where the branching times at the source and target of edge  $\mathbf{k}$  are  $t_I$  and  $t_{I'}$ .

As indicated by the  $\mathcal{T}$  subscript in (6.61), the total expression for  $P(T)$  includes terms additional to the contribution from the selected tree. In fact without additional subtraction terms (see [2, 5, 6, 18],) the model as formulated is not probability conserving. However, in phylogenetic inference (for example, maximum likelihood analyses) it is appropriate to



**Figure 2.** Binary labelling scheme for a tree on three leaves  $\mathcal{T} = (\vec{1}(\vec{2}\vec{4}))$  with branching events at intermediate times  $t_1, t_2$ . Nonzero rate constants for the model are chosen for the root and leaf momenta  $\vec{7} = (111), \vec{1} = (001), \vec{2} = (010)$  and  $\vec{4} = (100)$  respectively, together with a *single* additional momentum  $\vec{6} = (110)$ .

generate contributions from *all* candidate trees for unknown rates. In the present case, the additional terms arise of course from other admissible trees. In fact, even if only the rate constants for edges specific to a selected tree are nonzero, there are still contributions (in the operator approach) from intermediate states with non-propagating momenta, and these also arise in the combinatorics of the path integral perturbation expansion (see below). Thus in addition to (6.61) there are the trees with effective *trivalent* nodes,

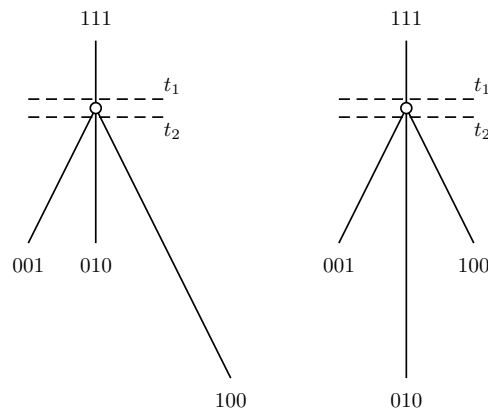
$$P_{\mathcal{T}_3} = \sum M_1^{-\alpha_1}_{\beta_1} M_2^{\alpha_2}_{\beta_2} M_4^{\prime\alpha_4}_{\beta_4} \Gamma^{\beta_1\beta_2\beta_4}_{\gamma_7} \cdot M_7^{\gamma_7}_{\beta_7} p^{\beta_7} \tag{6.62}$$

$$P_{\mathcal{T}_3} = \sum M_1^{-\alpha_1}_{\beta_1} M_2^{\prime\alpha_2}_{\beta_2} M_4^{\alpha_4}_{\beta_4} \Gamma^{\beta_1\beta_2\beta_4}_{\gamma_7} \cdot M_7^{\gamma_7}_{\beta_7} p^{\beta_7} \tag{6.63}$$

shown in figure 3. The *trivalency* comes by deleting edges for rates with  $R_{\mathbf{k}} = 0$  and re-joining the target and source nodes such that there is an effective 3-point vertex corresponding to a branching operator or an interaction vertex structure coefficient with components (compare (3.34))  $\Gamma_{\alpha}^{\beta\gamma\delta} = \delta_{\alpha}^{\beta}\delta_{\alpha}^{\gamma}\delta_{\alpha}^{\delta}$ . (Such an effective interaction term might also be viewed as the result of directly integrating out the  $\eta^{\mathbf{k}}, \eta^{\prime}_{\mathbf{k}}$  variables corresponding to  $R_{\mathbf{k}} = 0$ ). For the tree in question, the non-propagating momenta are  $\vec{3} = (011)$  and  $\vec{5} = (101)$  corresponding to the trees  $\mathcal{T}_3$  and  $\mathcal{T}_5$  respectively. The terms differ from one another because the edge evolution times  $T - t_1$  and  $T - t_2$  are distributed differently over the Markov matrices  $M_1, M_2$  and  $M_4$  as indicated by the ' in (6.62), (6.63) and the differing edge lengths in figure 3. Of course, it is always possible to regard these terms as vestigial contributions from standard *binary* trees with very short edges. In fact, since the usual counting relation between edges and leaves for binary trees obviously does not hold for the trivalent trees, the formal introduction of a scaling parameter would serve to distinguish these and similar noncanonical tree diagrams.

### 7. Conclusions and discussion

In this paper and the previous work [2], we have proposed a transcription of phylogenetic branching processes into the language of a stochastic dynamical system evolving according to an appropriate Markov rate operator on a suitably extended ‘state’ space. The analogy



**Figure 3.** Additional effective *non-binary* trees  $\mathcal{T}_3$  and  $\mathcal{T}_5$  contributing to the probability in the phylogenetic branching model for the three leaf case. Non-propagating momenta  $\vec{3} = (011)$  and  $\vec{5} = (101)$  produced by the branching interaction term at  $t_1$  cause effective trivalent vertices with different evolution times  $T - t_1, T - t_2$  on long and short leaf edges.

with statistical and particle physics is that the ‘particles’ in the phylogenetic context are the individual taxonomic units, and it is these which evolve in type and number (as in Markov models of reaction diffusion or birth–death processes, or in relativistic particle scattering) in the course of evolution. In [2] a conventional operator approach was taken, whereas in the present work the path integral formulation introduces to phylogenetics the familiar physical notions of ‘paths’ and ‘fields’ (over a discrete lattice). Our treatment including ‘interactions’ representing branching events, including explicit Feynman rules (table 1 in the appendix) establishes the equivalence of the path integral formulation to the operator version [2] via standard perturbation theory as the appropriate tool for completing the transcription.

The path integral language allows a range of techniques known in the context of the analysis of physical systems [5, 6, 18] to be deployed for phylogenetics. One immediate point is the relationship between the formulation of transition probabilities in ‘momentum’ space versus the dual ‘position’ space—standard in condensed matter systems, and also known in phylogenetics in the literature on transform techniques for phylogenetic inference involving the discrete Fourier Hadamard transformation, in principle to derive an edge rate spectrum for a phylogenetic tree directly from an observed data set of pattern probabilities [12, 19, 26]. In the present framework, momentum conservation is a reflection of translation invariance on the underlying discrete lattice.

General considerations for the path integral formulation include the further application of symmetry principles in various ways. For example, *continuous* Lie symmetry group actions on the path variables (phylogenetic fields), for example  $\bar{\eta} \rightarrow \bar{\eta} \cdot \bar{U}, \eta \rightarrow U \cdot \eta$  can be analysed for their effect on the dependence of the time evolution kernel on the various rate and time parameters of the model. This has been pursued in [23] (in the explicit tensor description) for the well-known Kimura 3ST model for four characters [15] where it was noted that the rate and branching operators intertwine the action of a  $U(1) \times U(1) \times U(1)$  (or  $\mathbb{C}^\times \times \mathbb{C}^\times \times \mathbb{C}^\times$ ) group so that the resultant group reduction from representations of  $SU(4)$  (or  $SL(4)$ ) is intimately related to the properties of this model (it is well known that the Kimura 3ST model and the related 2P model do belong to the class of *discrete* colour group models [12, 21]). More generally, the Lie symmetry approach allows rate models to be considered, in principle, in terms of a hierarchy of symmetry-breaking terms. For example, in molecular phylogenetics

at the protein mutation level, suitable symmetry groups would be those advocated recently in relation to the possible group-theoretical origins of the genetic code itself (see, for example, [1, 27]).

A deeper aspect of the branching model in the path integral formulation is the role of time reparametrizations,  $t \rightarrow \tau(t)$ , in connection with notions of the ‘molecular clock’. Given that

$$dt = d\tau \frac{dt}{d\tau}, \quad \delta(t - t_I) dt = \frac{\delta(\tau - \tau_I)}{|dt/d\tau|} \frac{dt}{d\tau} d\tau,$$

then clearly the evolution kernel has the following covariance property,

$$\mathcal{M}_T(t_I, R_{\mathbf{k}}(t)) = \mathcal{M}_T(\tau_I, R'_{\mathbf{k}}(\tau)), \quad \text{where } R'_{\mathbf{k}}(\tau) = R_{\mathbf{k}}(t) \cdot \frac{dt}{d\tau} \quad \text{and } \tau_I = \tau(t_I) \tag{7.64}$$

(it is assumed that  $dt/d\tau > 0$ , in particular  $\tau(t)$  is not orientation reversing). This is precisely the reason that, in standard probability approaches (see, for example, [12]), ‘dynamical’ considerations involving explicit time dependence can be absent—standard calculations require only the combinatorics of the tree (which is encoded in the present models via the branching times  $t_I$  and the choice of momenta for which rates are nonzero). However, as has been mentioned already, there is good reason to formulate the branching process temporally as presented here. In order for generalizations of the  $\delta$ -function forcing interaction terms to preserve the time reparametrization covariance (7.64), the introduction of an auxiliary phylogenetic ‘gauge’ field would be mandatory (as in some proper time formulations of relativistic field equations).

As an illustration of this dynamical perspective, suppose now that for some edge momentum  $\mathbf{k}^*$  the edge rate can be written as proportional to some standard rate matrix,

$$R_{\mathbf{k}^*}(t) = \lambda^*(t) R^*. \tag{7.65}$$

Then it is possible to define a phylogenetic ‘proper time’  $\tau^*$  (implicitly) as a function  $t$ , by solving the first order equation

$$\frac{dt}{d\tau^*} = \frac{1}{\lambda^*(t)}$$

together with some suitable initial condition, for example  $\tau^*_{t_I} = \tau^*(t_I) \equiv t_I$  where  $t_I$  is the branching time at the source node of edge  $\mathbf{k}^*$ . Then, with respect to this proper time, the edge rate  $R_{\mathbf{k}^*}(\tau^*)$  is by definition *constant*, and equal to  $R^*$ . By extension, if there exists a distinguished *tree path*  $\mathcal{P}^*$  from the root to some leaf node, along which *all* edge rate matrices possess the above multiplier property (7.65), a *global* phylogenetic proper time  $\tau^*$  exists for that tree path, with the rate matrices piecewise constant (constant along each edge). Finally, such a tree path phylogenetic proper time may always be adjusted to *coincide* with geological or archaeological time determinations at certain points by piecewise linear affine transformation(s) of the form  $\tau^* \rightarrow a\tau^* + b$  (which may be edge dependent along the distinguished path) without compromising the above arguments. An extreme example of this situation is of course the case of a stationary Markov process, wherein *each* rate matrix is (proportional to) a given fixed matrix  $R$ , and the (weighted) sum of elapsed times along each tree path from the root to a leaf node, is constant—in this case a molecular clock exists in the strongest sense. As usual however, it is still impossible to disentangle evolution occurring on some edges with standard strength for time  $\Delta t$ , from evolution occurring over time  $\lambda\Delta t$  with scaled rates  $(\lambda^{-1})R$ . In general, conclusions drawn

from studies of ‘time-dependent’ rate matrices should always be treated with caution because of reparametrization covariance. Related considerations for general Kolmogorov equations, related to non-stationary finite Markov processes, but without explicit recognition of the role of the group of time reparametrizations (diffeomorphisms), have been given in [20]; for a discussion of general time-dependent Markov processes, see [28]. The ‘intrinsic time’ of [20] is nothing but the above phylogenetic ‘proper time’  $\tau$ . This, in turn—interpreted as a gauge fixing choice—is essentially the Teichmüller parameter for the configuration space of an implicit ‘einbein’ path variable which carries gauge transformations associated with the group of time reparametrizations on the interval  $[0, T]$  (see, for example, [11]).

Within the present reformulation it is also possible to examine generalizations which may not be apparent in other contexts. An example would be analytical or at least systematic possibilities for the examination of the behaviour random trees in the limit of very large numbers of leaves, or of random branching events, for the purpose of comparative evolution studies. A further extension would be to include population processes such as mutation–selection effects into the models.

A final potentially important analytical tool is the fact that (as mentioned above) the closed form expression for the scattering probabilities represented by the evolution kernel generates contributions from *all* candidate trees for a given number of leaves. It is clear from our presentation that the characteristics of a specific tree can be encoded via the choice of nonzero rate constants for particular edge momenta, and that there may be several equivalent such assignments amongst the  $2^L$  admissible binary  $L$ -vectors. The exploitation of the interrelations of these assignments might give insights into the derivation of ‘invariants’ (in this case for the combinatorics of trees, rather than for differential topology, as in the case of topological quantum field theory) which could provide useful constraints in phylogenetic inference and maximum likelihood ‘optimal’ tree searches. Indeed, in maximum likelihood approaches themselves, it may be useful to have a formal representation of all contributing terms, without the need for explicit tree enumerations.

## Acknowledgments

PDJ and JGS thank the Department of Physics and Astronomy, and also the Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand, for hosting a visit in 2003 during which this work was presented. PDJ also acknowledges the hospitality of the Biosciences Research Centre, Massey University, during an earlier visit. This research was supported by the Australian Research Council grant no DP0344996. We thank the anonymous referees for suggestions for strengthening an earlier version.

## Appendix. Feynman rules for phylogenetic branching models

In this appendix, we develop systematic expansion methods in the form of Feynman rules, for the calculation of the time evolution of state probabilities in the model given by (5.53) and (5.59). This establishes that the model is formally equivalent to the standard prescription for calculating likelihood functions for phylogenetic trees, and provides the justification for the more qualitative discussion of the free and interacting cases given in sections 5 and 6 above.

Firstly note that the path integral representation of the (free) time evolution kernel  $\mathcal{M}_T^\circ(z, \zeta)$ , (4.46), (5.53) and (5.59) can be written in various equivalent symmetrized forms emphasizing the role of the boundary conditions, namely (using the generic form (4.46) to

suppress affixes)<sup>9</sup>,

$$\begin{aligned} \int_0^T dt (-i\eta'(t)\dot{\eta}(t) + R_t(i\eta', \eta)) + z\eta(T) &= \int_0^T dt (+i\dot{\eta}'(t)\eta(t) + R_t(i\eta', \eta)) + i\eta'(0)\zeta \\ &= \int_0^T dt \left( -i\frac{1}{2}(\eta'(t)\dot{\eta}(t) - \dot{\eta}'(t)\eta(t)) + R_t(i\eta', \eta) \right) + \frac{1}{2}z\eta(T) + \frac{1}{2}i\eta'(0)\zeta. \end{aligned} \tag{A.1}$$

Now consider the complete time evolution kernel extended by some ancillary path variables  $i\xi'(t), \xi(t)$ ,

$$\tilde{\mathcal{M}}_T := \int d[\eta'] d[\eta] \exp(i\eta' \cdot K \cdot \eta + i\xi' \cdot \eta + i\eta' \cdot \xi) \exp(+z\eta(T)) \exp S_1[i\eta', \eta], \tag{A.2}$$

such that  $\tilde{\mathcal{M}}_T \xrightarrow{\xi'=0=\xi} M_T$ . The notation ‘ $\cdot$ ’ in the exponential represents a definite integral of the occurring path variables with respect to time from  $t = 0$  to  $t = T$ , respecting of course the boundary conditions derived earlier, (4.47).<sup>10</sup> The notation ‘ $i\eta' \cdot K \cdot \eta$ ’ refers to the quadratic part of the integrand, in this case in the first of the forms (A.1). Finally, an additional (for the moment generic) ‘interaction’ term is included, with  $S_1$  being the integral of the normal kernel.

The aim is to consider the convolution of  $\tilde{\mathcal{M}}_T$  with the initial state probability generating function, in such a way that the expansion of the exponential of the interaction in a power series, together with the final state matrix element, and the folding with respect to the initial state probability tensor, are all reduced to formal derivatives with respect to the ancillary variables, acting on the expression for the ‘free’ kernel,

$$\tilde{\mathcal{M}}_T^\circ(z, \zeta) := \int d[\eta'] d[\eta] \exp(i\eta' \cdot K \cdot \eta + i\xi' \cdot \eta + i\eta' \cdot \xi) \exp(+z\eta(T)). \tag{A.3}$$

To this end, consider the complete generating function for the final probability state vector (compare (4.40)),

$$P_T(z) = \int d\zeta d\zeta' \tilde{\mathcal{M}}_T(z, \zeta) e^{-i\zeta'\zeta} P_0(\zeta') \Big|_{\xi=\xi'=0}. \tag{A.4}$$

The additional  $S_1[\eta, i\eta']$  interaction term in the exponential can be regarded, after a power series expansion, as a series of moments evaluated on the free kernel, so that

$$P_T(z) = e^{S_1[\frac{\partial}{\partial i\zeta'}, \frac{\partial}{\partial \zeta}]} \cdot \int d\zeta d\zeta' \tilde{\mathcal{M}}_T^\circ(z, \zeta) e^{-i\zeta'\zeta} P_0(\zeta') \Big|_{\xi=\xi'=0}. \tag{A.5}$$

Also if we are interested in a final state consisting of  $L$  taxonomic units, the relevant probability component is by definition the generating function derivative with respect to the appropriate

<sup>9</sup> Using either of the second two forms in the discussion following (5.53) leads to equivalent solutions, for example,

$$i\eta'_{k\alpha}(T) = i\eta'_{k\beta}(0)(M_{kT}^{-1})^\beta_\alpha, \quad \text{and then} \quad i\eta'(0)\zeta = i\eta'_{k\beta}(0)(M_{kT})^\beta_\alpha \zeta^\alpha$$

as before.

<sup>10</sup> Bearing in mind that the additional boundary contributions are for specific times, and are thus products, *not* integrals.



$z$  variables; for example,

$$P_T^{\alpha_1 \mathbf{k}_1 \dots \alpha_L \mathbf{k}_L} = \frac{\partial}{\partial z_{\alpha_1 \mathbf{k}_1}} \dots \frac{\partial}{\partial z_{\alpha_L \mathbf{k}_L}} P_T(z) \Big|_{z=0}. \tag{A.6}$$

From the dependence of the kernel on  $z$  (the first form in (A.1)), however, the derivatives merely bring down factors of  $\eta(T)$  with appropriate labels, which in turn are equivalent to the corresponding differentiations with respect to  $i\xi'$ :

$$P_T^{\alpha_1 \mathbf{k}_1 \dots \alpha_L \mathbf{k}_L} = \exp S_1 \left[ \frac{\partial}{\partial i\xi'}, \frac{\partial}{\partial \xi} \right] \frac{\partial}{\partial i\xi'_{\alpha_1 \mathbf{k}_1}} \dots \frac{\partial}{\partial i\xi'_{\alpha_L \mathbf{k}_L}} \cdot \int d\zeta d\zeta' \tilde{\mathcal{M}}_T(z, \zeta) e^{-i\zeta'\zeta} P_0(i\zeta') \Big|_{z=\xi=\xi'=0}. \tag{A.7}$$

Finally, from the *second* form of the kernel in (A.1), the path integral over  $\zeta$  will enforce a  $\delta$ -function constraint identifying  $i\zeta'$  with  $i\eta'(0)$ , or partial differentiation with respect to the appropriate components of  $\xi$ :

$$P_T^{\alpha_1 \mathbf{k}_1 \dots \alpha_L \mathbf{k}_L} = \exp S_1 \left[ \frac{\partial}{\partial i\xi'}, \frac{\partial}{\partial \xi} \right] \cdot \frac{\partial}{\partial i\xi'_{\alpha_1 \mathbf{k}_1}(T)} \cdot \frac{\partial}{\partial i\xi'_{\alpha_L \mathbf{k}_L}(T)} \cdot P_0 \left( \frac{\partial}{\partial \xi(0)} \right) \cdot \tilde{\mathcal{M}}_T^\circ(z, \zeta) \Big|_{z=\xi=\xi'=0}. \tag{A.8}$$

Turning to the evaluation of  $\tilde{\mathcal{M}}_T^\circ(z, \zeta)$  itself, note that the quadratic part of the integrand in (A.2) can be written as

$$\int \int_0^T dt dt' i\eta'(t') \cdot K(t, t') \eta(t) = \int \int_0^T dt dt' i\eta'(t) (-\partial_t \delta(t-t') + R\delta(t-t')) \eta(t'). \tag{A.9}$$

The formal completion of the square

$$i\eta' \cdot K \cdot \eta + i\xi' \cdot \eta + i\eta' \cdot \xi = i(\eta' + \xi' K^{-1}) \cdot K \cdot (\eta + K^{-1}\xi) - i\xi' \cdot K^{-1} \cdot \xi \tag{A.10}$$

suggests integrating out the resulting Gaussian after the change of variables  $i\eta' \rightarrow i(\eta' + \xi' K^{-1})$ ,  $\eta \rightarrow (\eta + K^{-1}\xi)$  (which has unit Jacobian), leaving the expression

$$\tilde{\mathcal{M}}_T^\circ(z, \zeta) = \exp(-i\xi' \cdot K^{-1} \cdot \xi) \tag{A.11}$$

up to normalization factors (including  $\det K^{-1}$ ) and boundary contributions. However, the explicit dependence on  $z$  and  $\zeta$  (which is to be integrated over in (A.4), (A.5)) has been circumvented by the device of formally introducing appropriate differentiations with respect to the  $\xi'$ ,  $\xi$  variables, so that (A.11) normalized with reference to the non-interacting case, is sufficient provided that  $K^{-1}$  is calculable. For the case of  $R$  constant this is easily checked to be

$$K^{-1}(t, t') = -\theta(t-t') e^{(t-t')R} \tag{A.12}$$

subject to  $K^{-1}(t, t') = 0$  if  $t \leq t'$ .

Consider then the non-interacting case (A.8), (A.12) with  $S_1 \equiv 0$ . Clearly, the necessity to set the ancillary variables equal to zero *after* differentiation means that the only viable initial probability state vector is one also with  $L$  extant taxa, *and* with identical momentum labels.

Explicitly, we have

$$P_T^{\alpha_1 \mathbf{k}_1 \dots \alpha_L \mathbf{k}_L} = \frac{\partial}{\partial i \xi'_{\alpha_1 \mathbf{k}_1}(T)} \dots \frac{\partial}{\partial i \xi'_{\alpha_L \mathbf{k}_L}(T)} P_0^{\beta_1 \mathbf{k}_1 \dots \beta_L \mathbf{k}_L} \frac{\partial}{\partial \xi^{\beta_1 \mathbf{k}_1}(0)} \dots \frac{\partial}{\partial \xi^{\beta_L \mathbf{k}_L}(0)} \times \exp \int \int_0^T dt dt' \theta(t - t') \sum_{\gamma, \delta, \mathbf{m}} i \xi'_{\gamma \mathbf{m}}(t) M_{(t-t') \mathbf{m} \delta}^\gamma \xi^{\delta \mathbf{m}}(t') \Big|_{z=\xi=\xi'=0}. \quad (\text{A.13})$$

Differentiations with respect to  $i \xi'$ ,  $\xi$  with the corresponding momentum labels must be paired, leading finally to

$$P_T^{\alpha_1 \mathbf{k}_1 \dots \alpha_L \mathbf{k}_L} = \sum_{\delta_i} M_{T \mathbf{k}_1}^{\alpha_1 \beta_1} M_{T \mathbf{k}_2}^{\alpha_2 \beta_2} \dots M_{T \mathbf{k}_L}^{\alpha_L \beta_L} P_0^{\beta_1 \mathbf{k}_1 \dots \beta_L \mathbf{k}_L} \quad (\text{A.14})$$

as was derived informally in (5.52), (5.58).

Turning to the interacting case, we are interested in the final state probability for  $L$  taxonomic units, assigned momenta  $\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_L$  say, arising from an initial state with one taxon (the root) with momentum  $\mathbf{k}_0$ , thus the probability component for an  $L \leftarrow 1$  scattering process in the model. Once again, the necessity to set the ancillary variables equal to zero *after* differentiation selects nonvanishing contributions corresponding to precisely degree  $L - 1$  in the power series expansion of the exponential of the interaction term  $S_1[\frac{\partial}{\partial i \xi'}, \frac{\partial}{\partial \xi}]$  (see (A.8) and (6.60)):

$$P_T^{\alpha_1 \mathbf{k}_1 \dots \alpha_L \mathbf{k}_L} = \frac{1}{(L - 1)!} \left[ - \int_0^T dt \frac{(-i)^3}{2} \sum_I \sum_{\mathbf{k}, \mathbf{l}, \mathbf{m}} \delta(t - t_I) \delta(\mathbf{k} - \mathbf{l} - \mathbf{m}) \frac{\partial}{\partial \xi^{\mathbf{l} \alpha}(t)} \times \frac{\partial}{\partial \xi^{\mathbf{m} \beta}(t)} \Gamma^{\alpha \beta \gamma} \frac{\partial}{\partial \xi'_{\mathbf{m} \gamma}(t)} \right]^{(L-1)} \cdot \frac{\partial}{\partial i \xi'_{\alpha_1 \mathbf{k}_1}(T)} \dots \frac{\partial}{\partial i \xi'_{\alpha_L \mathbf{k}_L}(T)} \cdot P_0^{\beta_0 \mathbf{k}_0} \frac{\partial}{\partial \xi^{\beta_0 \mathbf{k}_0}(0)} \cdot \exp - \left[ \int \int_0^T dt dt' \theta(t - t') \sum_{\gamma, \delta, \mathbf{m}} i \xi'_{\gamma \mathbf{m}}(t) M_{(t-t') \mathbf{m} \delta}^\gamma \xi^{\delta \mathbf{m}}(t') \right] \Big|_{z=\xi=\xi'=0}. \quad (\text{A.15})$$

It is convenient at this stage also to choose canonical momenta (binary  $L$ -vectors, with a scaling of  $\pi$  understood)  $\mathbf{k}_0 = (1, 1, \dots, 1)$  for the root, and  $\mathbf{k}_1 = (0, 0, \dots, 1), \mathbf{k}_2 = (0, \dots, 1, 0), \dots, \mathbf{k}_L = (0, \dots, 0, 1)$  for the edges (or decimal equivalents  $\bar{1}, \bar{2}, \bar{4}, \dots$ ). For formal analysis with a specific tree, it may in fact be more powerful to consider *all* such admissible  $L \leftarrow 1$  momentum routing schemes; however, for combinatorial purposes any fixed assignment is sufficient.

For  $L = 2$  there is only one interaction, whose time is forced to be  $t = t_1$ . Performing the differentiation of the exponential of the free kernel with respect to  $\xi^{\beta_0 \mathbf{k}_0}(0)$  gives

$$P_T^{\alpha_1 \mathbf{k}_1 \alpha_2 \mathbf{k}_2} = \left[ + \frac{1}{2} \sum_{\mathbf{k}, \mathbf{l}, \mathbf{m}} \delta(\mathbf{k} - \mathbf{l} - \mathbf{m}) \frac{\partial}{\partial \xi^{\mathbf{l} \alpha}(t_1)} \frac{\partial}{\partial \xi^{\mathbf{m} \beta}(t_1)} \Gamma^{\alpha \beta \gamma} \frac{\partial}{\partial i \xi'_{\mathbf{k} \gamma}(t_1)} \right] \cdot \frac{\partial}{\partial i \xi'_{\alpha_1 \mathbf{k}_1}(T)} \times \frac{\partial}{\partial i \xi'_{\alpha_2 \mathbf{k}_2}(T)} \cdot \left[ + \int_0^T dt i \xi'_{\lambda \mathbf{k}_0}(t) M_{t \mathbf{k}_0}^{\lambda \beta_0} \right] P_0^{\beta_0 \mathbf{k}_0} \cdot \exp + \left[ \int \int_0^T dt dt' \theta(t - t') \sum_{\gamma, \delta, \mathbf{m}} i \xi'_{\rho \mathbf{m}}(t) (M_{(t-t') \mathbf{m}}^\rho)_\sigma \xi^{\sigma \mathbf{m}}(t') \right] \Big|_{\xi=0 \equiv \xi'}$$

and carrying out the two remaining  $\xi$  differentiations leads to

$$P_T^{\alpha_1 \mathbf{k}_1 \alpha_2 \mathbf{k}_2} = \left[ + \frac{1}{2} \sum_{\mathbf{k}, \mathbf{l}, \mathbf{m}} \delta(\mathbf{k} - \mathbf{l} - \mathbf{m}) \Gamma^{\alpha\beta\gamma} \frac{\partial}{\partial i \xi'_{\alpha_1 \mathbf{k}_1}(T)} \frac{\partial}{\partial i \xi'_{\alpha_2 \mathbf{k}_2}(T)} \frac{\partial}{\partial i \xi'_{\mathbf{k}_\gamma}(t_1)} \right] \\ \cdot \left[ + \int_{t_1}^T dt i \xi'_{\lambda_1}(t) M_{t_1}^{\lambda \alpha} \right] \left[ + \int_{t_1}^T dt i \xi'_{\mu \mathbf{m}}(t) M_{t \mathbf{m}}^{\mu \beta} \right] \left[ + \int_0^T dt i \xi'_{\nu \mathbf{k}_0}(t) M_{t \mathbf{k}_0}^{\nu \beta_0} \right] \\ \times P_0^{\beta_0 \mathbf{k}_0} \cdot \exp + \left[ \int_0^T \int_0^T dt dt' \theta(t - t') \sum_{\gamma, \delta, \mathbf{m}} i \xi'_{\rho \mathbf{m}}(t) M_{(t-t') \mathbf{m}}^{\rho \sigma} \xi^{\sigma \mathbf{m}}(t') \right] \Bigg|_{\xi=0 \equiv \xi'}$$

For a nonzero result, the remaining  $\xi'(t_1)$  and two  $\xi'(T)$  differentiations can only be applied to the terms standing in front of the exponential. Moreover, the implicit  $\theta$  terms require the  $\xi'(t_1)$  differentiation to be applied only to the  $\mathbf{k}_0$  integral, thus fixing  $\mathbf{k} = \mathbf{k}_0$ . Finally, since  $\mathbf{k}_0 = \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{l} + \mathbf{m}$  there are two equivalent ways to apply the remaining differentiations (cancelling the symmetry factor  $\frac{1}{2}$  in the interaction term) giving finally

$$P_T^{\alpha_1 \mathbf{k}_1 \alpha_2 \mathbf{k}_2} = \Gamma^{\alpha\beta\gamma} (M_{(T-t_1) \mathbf{k}_1})^{\alpha_1 \alpha} (M_{(T-t_1) \mathbf{k}_2})^{\alpha_2 \beta} (M_{t_1 \mathbf{k}_0})^{\gamma \beta_0} P_0^{\beta_0 \mathbf{k}_0}. \quad (\text{A.16})$$

In the general case, systematic diagrammatical rules (Feynman rules) can easily be ascribed and tabulated for the evaluation of (A.15). On the basis of the above  $L = 2$  (first order) case and similar considerations for  $L = 3$  (second order), all possible probability component contributions for  $L$  taxa are constructed from the matrix element for  $L \leftarrow 1$  scattering as follows.

#### *Feynman rules for phylogenetic trees*

- (1) Diagrams consist of  $2L - 1$  directed edges,  $L - 1$  vertices with internal nodes, one external root and  $L$  leaf nodes.
- (2) To each element is assigned character and momentum labels as in table 1.
- (3) Specifically, root and leaf edge momenta are assigned canonical binary  $L$ -vectors (see the text); momentum conservation between ingoing and outgoing edge momenta is imposed.
- (4) Vertices (internal nodes) are assigned interaction times ordered  $t_1 < t_2 < \dots < t_{L-1}$ .
- (5) The root node is assigned time  $t = 0 = t_0$ , and the leaves are assigned time  $t = T = t_L$ .

To these labelled diagrammatical elements, the following algebraic terms are associated:

- (1) For each directed edge, a Markov transition matrix for time interval  $\Delta = (t_{I'} - t_I)$ ,  $0 \leq I \leq L - 1$  between the target and source nodes, and for its assigned edge momentum, and matrix element determined by the source and target character labels assigned (see table 1).
- (2) To each vertex, an appropriate component of the  $\Gamma$  structure coefficient (see table 1).
- (3) Consistent combinations of these elements, summed over all internal momenta and character indices, with valid momentum conservation, correspond to contributions from all possible labelled  $L$  leaf binary trees.

Using these rules, likelihoods can thus be written down autonomously and diagrammatically, without reference to the path integral context; however, as stressed in the main text, the utility of the path integral formulation is precisely to provide a self-contained prescription for them without the explicit need to enumerate trees.

## References

- [1] Bashford J D and Jarvis P D 2000 The genetic code as a periodic table: algebraic aspects *Biosystems* **57** 147–61
- [2] Bashford J D and Jarvis P D 2001 Quantum field theory and phylogenetic branching *J. Phys. A: Math. Gen.* **34** L703–L707
- [3] Chang J T 1996 Full reconstruction of Markov models on evolutionary trees: identifiability and consistency *Math. Biosci.* **137** 51–73
- [4] Diaconis P 1985 *Group Representations in Probability and Statistics (Lecture Notes—Monograph Series vol 11)* ed S Shastri *et al* (Cambridge, MA: Institute of Mathematical Statistics, Harvard University)
- [5] Doi M 1976 Second quantization representation for classical many-particle system *J. Phys. A: Math. Gen.* **9** 1465–77
- [6] Doi M 1976 Stochastic theory of diffusion-controlled reaction *J. Phys. A: Math. Gen.* **9** 1479–95
- [7] Farris J S 1973 A probability model for inferring evolutionary trees *Syst. Zool.* **22** 250–6
- [8] Felsenstein J 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach *J. Mol. Evol.* **17** 368–76
- [9] Felsenstein J 2004 *Inferring Phylogenies* (Sunderland, MA: Sinauer Associates)
- [10] Jarvis P D, Sumner J G and Bashford J D 2005 Markov group invariants, plethysms and phylogenetic branching in preparation
- [11] Govaerts J 1991 *Hamiltonian Quantisation and Constrained Dynamics (Leuven Notes in Mathematical and Theoretical Physics, Series B, Theoretical Particle Physics vol 4)* (Leuven: Leuven University Press)
- [12] Steel M, Hendy M D and Penny D 1998 Reconstructing phylogenies from nucleotide pattern probabilities: a survey and some new results *Discrete Appl. Math.* **88** 367–96
- [13] Alcaraz F C, Droz M, Henkel M and Rittenberg V 1994 Reaction–diffusion processes, critical dynamics, and quantum chains *Ann. Phys., NY* **230** 250–302
- [14] Johnson J E 1985 Markov-type Lie groups in  $GL(n, R)$  *J. Math. Phys.* **26** 252–7
- [15] Kimura M 1981 Estimation of evolutionary distances between homologous nucleotide sequences *Proc. Natl Acad. Sci. USA* **78** 454–8
- [16] Rodriguez F, Oliver J L, Marin A and Medina J R 1990 The general stochastic model of nucleotide substitution *J. Theor. Biol.* **142** 485–501
- [17] McCullagh P 1987 *Tensor Methods in Statistics* (London: Chapman and Hall)
- [18] Peliti L 1985 Path integral approach to birth–death processes on a lattice *J. Physique* **46** 1469–83
- [19] Hendy M D, Penny D and Steel M A 1994 Discrete Fourier analysis for evolutionary trees *Proc. Natl Acad. Sci. USA* 3339–43
- [20] Goodman Gerald S 1970 An intrinsic time for non-stationary finite Markov chains *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **16** 165–80
- [21] Semple C and Steel M 2003 *Phylogenetics* (Oxford: Oxford University Press)
- [22] Sudbery A 1986 *Quantum Mechanics and the Particles of Nature: an Outline for Mathematicians* (Cambridge: Cambridge University Press)
- [23] Bashford J D, Jarvis P D, Sumner J G and Steel M A 2004  $U(1) \times U(1) \times U(1)$  symmetry of the Kimura 3ST model and phylogenetic branching processes *J. Phys. A: Math. Gen.* **37** L1–L9
- [24] Sumner J G and Jarvis P D 2004 Entanglement invariants and phylogenetic branching *J. Math. Biol.* **51** 18–36 (Preprint q-bio.PE/0402007)
- [25] Sumner J G and Jarvis P D 2005 Using the tangle: an improved method for construction of phylogenetic distance matrices, in preparation
- [26] Székely L A, Erdős P L, Steel M A and Penny D 1993 A Fourier inversion formula for evolutionary trees *Appl. Math. Lett.* **6** 13–6
- [27] Bashford J D, Tsohantjis I and Jarvis P D 1998 A supersymmetric model for the evolution of the genetic code *Proc. Natl Acad. Sci. USA* **95** 987–92
- [28] Rindos A, Woollet S, Viniotis I and Trivedi K 1995 Exact methods for the transient analysis of nonhomogeneous continuous time Markov chains *Second International Workshop of Markov Chains* ed W J Stewart (Dordrecht/New York: Kluwer/Academic)