*The Annals of Applied Probability*
2007, Vol. 17, No. 4, 1172–1201
DOI: 10.1214/105051607000000069


# EXCHANGEABLE PARTITIONS DERIVED FROM MARKOVIAN COALESCENTS

BY RUI DONG,[1] ALEXANDER GNEDIN AND JIM PITMAN[2]

*University of California, Berkeley, Utrecht University and University of California, Berkeley*

Kingman derived the Ewens sampling formula for random partitions describing the genetic variation in a neutral mutation model defined by a Poisson process of mutations along lines of descent governed by a simple coalescent process and observed that similar methods could be applied to more complex models. Möhle described the recursion which determines the generalization of the Ewens sampling formula in the situation where the lines of descent are governed by a $\Lambda$-coalescent, which allows multiple mergers. Here, we show that the basic integral representation of transition rates for the $\Lambda$-coalescent is forced by sampling consistency under more general assumptions on the coalescent process. Exploiting an analogy with the theory of regenerative partition structures, we provide various characterizations of the associated partition structures in terms of discrete-time Markov chains.

**1. Introduction.** The theory of random coalescent processes starts with Kingman's series of papers [20–22] in 1982. The idea comes from biological studies for genealogy of haploid model [5]: given a large population with many generations, you track backward in time the family history of each individual in the current generation. As you track further, the family lines coalesce with each other, eventually all terminating at a common ancestor of current generation. The same mathematical process may be interpreted in another way as describing collisions of an aggregating system of physical particles. In Kingman's coalescent process [20], each collision only involves two parts. This idea is extended to coalescent with multiple collisions in [29, 30], where every collision can involve two or more parts. This model is further developed into the theory of coalescent with simultaneous multiple collisions in [25, 33]. See [3, 4, 6, 9, 31, 32, 34, 36] for related developments.

Kingman [22] indicated a basic connection between random partitions of natural interest in genetics, and coalescent processes. Suppose, in the haploid case, the family line of current generation is modeled by Kingman's coalescent and the

Received March 2006; revised December 2006.

[1]Supported by NSF Grant DMS-04-05779.

[2]Supported in part by NSF Grant DMS-04-05779.

*AMS 2000 subject classifications.* Primary 60G09; secondary 60C05.

*Key words and phrases.* Exchangeable partitions, $\Lambda$-coalescent with freeze, consistency, decrement matrix.

mutations are applied along the family lines by using a Poisson process with rate $\theta/2$ for some nonnegative number $\theta$. Define a partition by saying that two individuals are in the same block if there is no mutation along their family lines before they coalesce. Then the resulting random partition is governed by the Ewens sampling formula with parameter $\theta$; see [28], Section 5.1, Exercise 2, and [2, 27] for a review and for more on this idea. Recently, Möhle [23] applied this idea to the genealogy tree modeled by coalescents with multiple collisions and simultaneous multiple collisions. He studied the resulting family of partitions and derived a recursion which determines them. In [24], Möhle showed that the partition derived from coalescent with multiple collisions is regenerative in the sense of [13–15] if and only if the underlying coalescent is Kingman's coalescent or a hook case, corresponding to the extreme cases when the characterization measure $\Lambda$ of coalescent with multiple collisions concentrates at 0 or 1, respectively. In particular, the intersection of Möhle's family of partitions with Pitman's two-parameter family is the one-parameter Ewens' family.

Here, we offer a different approach to the family of random partitions generated by Poisson marking along the lines of descent of a $\Lambda$-coalescent. We study partitions with an additional feature, assigning each part one of two possible states: *active* or *frozen*. We introduce a new class of continuous-time partition-valued coalescent processes called *coalescents with freeze*, which are characterized by an underlying measure determining collision rates, together with a freezing rate. Every coalescent with freeze has a terminal state with all blocks frozen, called the *final partition* of this process, whose distribution is characterized by the recursion of Möhle [23]. In the spirit of [15, 13], we focus here on the discrete-time chains embedded in the coalescent with freeze and from the consistency of their transition operators, we derive a backward recursion satisfied by the decrement matrix, analogous to [15], Theorem 3.3. This decrement matrix determines the partition through Möhle's recursion. As in [15], we use algebraic methods to derive an integral representation for the decrement matrix. Also, adapting an idea from [13], we establish a uniqueness result by constructing another Markov chain, with state space the set of partitions of a finite set, whose unique stationary distribution is the law of the final partition restricted to this set. We analyze in detail the case of coalescent with freeze when no simultaneous multiple collisions are permitted, leaving the more general case to another paper.

The remainder of the paper is organized as follows. Some notation and background are introduced in Section 2, together with a review of Möhle's result. In Section 3, the coalescent with freeze is defined and the relation between our method and Möhle's method is discussed. In Section 4, we detail the study of coalescent with freeze in terms of the freeze-and-merge (FM) operators of the embedded finite discrete chain, whose consistency with sampling derives a backward recursion for the decrement matrix. In Section 5, the Markov chain with sample-and-add (SA) operation is introduced and the law of the partition in our study is identified as the unique stationary distribution of this chain. In Section 6, we derive the integral representation for an infinite decrement matrix. This gives another approach to Möhle's partitions via consistent freeze-and-merge chains, which may be seen as discrete-time jumping processes associated with the $\Lambda$-coalescent with freeze. Section 7 provides an alternate approach to the representation of an infinite decrement matrix in terms of a positivity condition on a single sequence. Section 8 offers some results about the structure of the random set of freezing times derived from a coalescent with freeze. Finally, in Section 9, we point out some striking parallels with our previous work on regenerative partition structures which guided this study.

## 2. Some notation and background.

Following the notation of [28], for any finite set $F$, a partition of $F$ into $\ell$ blocks, also called a *finite set partition*, is an unordered collection of nonempty disjoint sets $\{A_1, \dots, A_\ell\}$ whose union is $F$. In particular, we consider partitions of the set $[n] := \{1, 2, \dots, n\}$ for $n \in \mathbb{N}$. We use $\mathcal{P}_{[n]}$ to denote the set of all partitions of $[n]$. A *composition* of the positive integer $n$ is an ordered sequence of positive integers $(n_1, n_2, \dots, n_\ell)$ with $\sum_{i=1}^{\ell} n_i = n$, where $\ell \in \mathbb{N}$ is the number of parts. We use $\mathcal{C}_n$ to denote the set of all compositions of $n$ and $\mathcal{P}_n$ to denote the set of nonincreasing compositions of $n$, also called *partitions of* $n$.

Let $\pi_n = \{A_1, A_2, \dots, A_\ell\}$ denote a generic partition of $[n]$; we may write $\pi_n \vdash [n]$ to indicate this fact. The *shape function* from partitions of the set $[n]$ to partitions of the positive integer $n$ is defined by

$$\text{shape}(\pi_n) = (|A_1|, |A_2|, \dots, |A_\ell|)^{\downarrow}, \tag{2.1}$$

where $|A_i|$ is the size of block $A_i$ which represents the number of elements in the block and "$\downarrow$" means arranging the sequence of sizes in nonincreasing order.

A random partition $\Pi_n$ of $[n]$ is a random variable taking values in $\mathcal{P}_{[n]}$. It is called *exchangeable* if its distribution is invariant under the action on partitions of $[n]$ by the symmetric group of permutations of $[n]$. Equivalently, the distribution of $\Pi_n$ is given by the formula

$$\mathbb{P}(\Pi_n = \{A_1, A_2, \dots, A_\ell\}) = p_n(|A_1|, |A_2|, \dots, |A_\ell|) \tag{2.2}$$

for some symmetric function $p_n$ of compositions of $n$. We call $p_n$ the *exchangeable partition probability function* (*EPPF*) of $\Pi_n$.

An *exchangeable random partition of* $\mathbb{N}$ is a sequence of exchangeable set partitions $\Pi_\infty = (\Pi_n)_{n=1}^{\infty}$ with $\Pi_n \vdash [n]$, subject to the *consistency condition*

$$\Pi_n|_m = \Pi_m, \tag{2.3}$$

where the restriction operator $|_m$ acts on $\mathcal{P}_{[n]}$, $n > m$, by deleting elements $m+1, m+2, \dots, n$. The distribution of such an exchangeable random partition of $\mathbb{N}$ is determined by the function $p$ defined on the set of all integer compositions

$\mathcal{C}_\infty := \bigcup_{i=1}^{\infty} \mathcal{C}_i$, which coincides with the EPPF $p_n$ of $\Pi_n$ when acting on $\mathcal{C}_n$. This function $p$ is called the *infinite EPPF associated with* $\Pi_\infty = (\Pi_n)_{n=1}^{\infty}$. The consistency condition (2.3) translates into the following *addition rule* for the EPPF $p$: for each positive integer $n$ and each composition $(n_1, n_2, \ldots, n_\ell)$ of $n$,

$$p(n_1, n_2, \ldots, n_\ell) = p(n_1, n_2, \ldots, n_\ell, 1) + \sum_{i=1}^{\ell} p(n_1, \ldots, n_i + 1, \ldots, n_\ell), \tag{2.4}$$

where $(n_1, \ldots, n_i + 1, \ldots, n_\ell)$ is formed from $(n_1, \ldots, n_\ell)$ by adding 1 to $n_i$. Conversely, if a nonnegative function $p$ on compositions satisfies (2.4) and the normalization condition $p(1) = 1$, then by Kolmogorov's extension theorem, there exists an exchangeable random partition $\Pi_\infty$ with EPPF $p$.

Similar definitions apply to a finite sequence of consistent exchangeable random set partitions $(\Pi_m)_{m=1}^{n}$ with $\Pi_m \vdash [m]$, where $n$ is some fixed positive integer. The *finite* EPPF $p$ of such a sequence can be defined as the unique recursive extension of $p_n$ by the addition rule (2.4) to all compositions $(n_1, n_2, \ldots, n_\ell)$ of $m < n$.

Let $\mathcal{P}_\infty$ denote the set of all partitions of $\mathbb{N}$. We identify each $\pi_\infty \in \mathcal{P}_\infty$ as the sequence $(\pi_1, \pi_2, \ldots) \in \mathcal{P}_{[1]} \times \mathcal{P}_{[2]} \times \cdots$, where $\pi_n = \pi_\infty|_n$ is the restriction of $\pi_\infty$ to $[n]$ obtained by deleting all elements greater than $n$. Endow $\mathcal{P}_\infty$ with the topology it inherits as a subset of $\mathcal{P}_{[1]} \times \mathcal{P}_{[2]} \times \cdots$ with the product of discrete topologies, so the space $\mathcal{P}_\infty$ is compact and metrizable. Following [9, 20, 29], call a $\mathcal{P}_\infty$-valued stochastic process $(\Pi_\infty(t), t \geq 0)$ a *coalescent* if it has càdlàg paths and $\Pi_\infty(s)$ is a *refinement* of $\Pi_\infty(t)$ for every $s < t$. For a nonnegative finite measure $\Lambda$ on the Borel subsets of [0, 1], a $\Lambda$-*coalescent* is a $\mathcal{P}_\infty$-valued Markov coalescent $(\Pi_\infty(t), t \geq 0)$ whose restriction $(\Pi_n(t), t \geq 0)$ to $[n]$ is for each $n$ a Markov chain such that when $\Pi_n(t)$ has $b$ blocks, each $k$-tuple of blocks of $\Pi_n(t)$ is merging to form a single block at rate $\lambda_{b,k}$, where

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k} \Lambda(dx) \qquad (2 \leq k \leq b < \infty). \tag{2.5}$$

The measure $\Lambda$ which characterizes the coalescent is derived from the consistency requirement, that is, for any positive integers $0 < m < n < \infty$ and $\pi_n \vdash [n]$, the restricted process $(\Pi_n(t)|_m, t \geq 0)$ given $\Pi_n(0) = \pi_n$ has the same law as $(\Pi_m(t), t \geq 0)$ given $\Pi_m(0) = \pi_n|_m$. This condition is fulfilled if and only if the array of rates $(\lambda_{b,k})$ satisfies

$$\lambda_{b,k} = \lambda_{b+1,k} + \lambda_{b+1,k+1} \qquad (2 \leq k \leq b < \infty). \tag{2.6}$$

The integral representation (2.5) can be derived from (2.6) via de Finetti's Theorem [29], Lemma 18.

When $\Lambda = \delta_0$, this reduces to Kingman's coalescent [20–22] with only binary merges. When $\Lambda$ is the uniform distribution on [0, 1], the coalescent is the Bolthausen–Sznitman coalescent [4]. In [33], this construction is further developed to build the $\Xi$-coalescent, where the measure $\Xi$ on infinite simplex characterizes the rates of simultaneous multiple collisions.

Möhle [23] studied the following generalization of Kingman's model [22]. Take a genetic sample of $n$ individuals from a large population and label them as $\{1, 2, \ldots, n\}$. Suppose that the ancestral lines of these $n$ individuals evolve by the rules of a $\Lambda$-coalescent and that given the genealogical tree, whose branches are the ancestral lines of these individuals, mutations occur along the ancestral lines according to a Poisson point process with rate $\rho > 0$. The infinitely-many-alleles model is assumed, which means that when a gene mutates, a brand new type appears. Define a random partition of $[n]$ by declaring individuals $i$ and $j$ to be in the same block if and only if they are of the same type, that is, if either $i = j$ or there are no mutations along the ancestral lines of $i$ and $j$ before these lines coalesce. These random partitions are exchangeable, and consistent as $n$ varies. The EPPF of this random partition is the unique solution $p$ with $p(1) = 1$ of *Möhle's recursion*: for each positive integer $n$ and each composition $(n_1, n_2, \ldots, n_\ell)$ of $n$,

$$\begin{aligned} p(n_1, n_2, \ldots, n_\ell) = {} & \frac{q(n:1)}{n} \sum_{j:n_j=1} p(\ldots, \widehat{n_j}, \ldots) \\ & + \sum_{k=2}^{n} q(n:k) \sum_{j:n_j \geq k} \frac{\binom{n_j}{k}}{\binom{n}{k}} p(\ldots, n_j - k + 1, \ldots), \end{aligned} \tag{2.7}$$

where $(\ldots, \widehat{n_j}, \ldots)$ is formed from $(n_1, n_2, \ldots, n_\ell)$ by removing part $n_j$, $(\ldots, n_j - k + 1, \ldots)$ is formed from $(n_1, n_2, \ldots, n_\ell)$ by only changing $n_j$ to $n_j - k + 1$ and $q(b:k)$ is the stochastic matrix

$$q(b:k) = \frac{\Phi(b:k)}{\Phi(b)} \qquad (1 \leq k \leq b \leq n), \tag{2.8}$$

where

$$\Phi(b:1) = \rho b, \tag{2.9}$$

$$\Phi(b:k) = \binom{b}{k} \lambda_{b,k} = \binom{b}{k} \int_0^1 x^{k-2}(1-x)^{b-k} \Lambda(dx) \qquad (2 \leq k \leq b), \tag{2.10}$$

$$\Phi(b) = \sum_{k=1}^{b} \Phi(b:k) = \int_0^1 \frac{1-(1-x)^b - bx(1-x)^{b-1}}{x^2} \Lambda(dx) + \rho b. \tag{2.11}$$

If, at some time $t \geq 0$, there are exactly $b$ lines of descent whose associated genealogical trees of depth $t$ contain no mutations, then $\Phi(b:1)$ is the total rate of mutations along one of these $b$ lines, $\Phi(b:k)$ is the total rate of $k$-fold merges among these lines and $\Phi(b)$ is the total rate of events of either kind.

Möhle [23] derived the recursion (2.7) by conditioning on whether the first event met tracing back in time from the current generation is a mutation or collision. On the left-hand side of (2.7), $p(n_1, n_2, \ldots, n_\ell)$ is the probability of ending up with any *particular* partition $\pi_n$ of the set $[n]$ into $\ell$ blocks of sizes $(n_1, n_2, \dots, n_\ell)$. On the right-hand side, $q(n:1)$ is the chance that starting from the current generation, one of the $n$ genes mutates before any collision; for this to happen together with the specified partition of $[n]$, the individual with this gene must be chosen from those among the singletons of $\pi_n$, with chance $1/n$ for each different choice, and after that, the restriction of the coalescent process to a subset of $[n]$ of size $n-1$ must end up generating the restriction of $\pi_n$ to that set. Similarly, $q(n:k)$ is the chance that the first event met is $k$ out of $n$ genes coalescing to the same block. Again, the $k$ individuals bearing these $k$ genes must be chosen from a block of $\pi_n$ of size $n_j \geq k$, so the chance for possible choices from a block with size $n_j$ is $\binom{n_j}{k}/\binom{n}{k}$ and given exactly which $k$ individuals are chosen, the restriction of the coalescent process to some set of $n-k+1$ lines of descent must end up generating a particular partition of these $n-k+1$ lines into sets of sizes $(\dots, n_j - k + 1, \dots)$. The multiplication of various probabilities is justified by the strong Markov property of the $\Lambda$-coalescent at the time of the first event and by the special symmetry property that lines of descent representing blocks of individuals coalesce according to the same dynamics as if they were singletons.

In this paper, we step back from these detailed dynamics of the $\Lambda$-coalescent with mutations to consider the following questions related to Möhle's recursion (2.7) and associated partition-valued processes. We choose to ignore the special form (2.8) of the matrix $(q(n:k); 1 \leq k \leq n < \infty)$ derived from the $(\Lambda, \rho)$ and analyze Möhle's recursion (2.7) as an abstract relation between a stochastic matrix $q$ and a function of compositions $p$. In particular, we ask the following questions:

1. For which probability distributions $q(n:k)$, $1 \leq k \leq n$, on $[n]$ is Möhle's recursion (2.7) satisfied by the EPPF $p$ of some exchangeable random partition of $[n]$ and is this $p$ uniquely determined?
2. How can such random partitions be characterized probabilistically?
3. Can such random partitions of $[n]$ be consistent as $n$ varies for any $q$ other than that derived from $(\Lambda, \rho)$ as above?

We stress that in the first two questions, the recursion (2.7) is only required to hold for a single value of $n$, while in the third question, (2.7) must hold for all $n = 1, 2, \dots$. The answer to the first question is that for each fixed probability distribution $q(n:k)$, $1 \leq k \leq n$, on $[n]$, Möhle's recursion (2.7) determines a unique EPPF $p$ for an exchangeable random partition of $[n]$ (Theorem 4.6). Answering the second question, we characterize the distribution of this random partition in two different ways: first, as the terminal state of a discrete-time Markovian coalescent process, the *freeze-and-merge chain* introduced in Section 4, and second, as the stationary distribution of a partition-valued Markov chain with quite a different transition mechanism, the *sample-and-add chain* introduced in Section 5. The answer to the third question is positive if we restrict $n$ to some bounded range of values, for some, but not all, $q$ (see Section 4), but negative if we require consistency for all $n$ (Theorem 6.2): if an infinite EPPF $p$ solves Möhle's recursion (2.7) for all $n$ for some triangular matrix $q$ with nonnegative entries, then $q$ must have the form (2.8) for some $(\Lambda, \rho)$.

We were guided in this analysis by a remarkable parallel between this theory of finite and infinite partitions subject to Möhle's recursion (2.7) and the theory of *regenerative partitions* developed in [13–15]. Many of these parallels are summarized in Section 9.

There is an important distinction between the recursions (2.4) and (2.7). The recursion (2.4) has many solutions since it is a *backward* recursion, from larger values of $n$ to smaller. By contrast, (2.7) is a *forward* recursion, from smaller values of $n$ to larger. Consequently it is obvious that given an arbitrary infinite triangular stochastic matrix $q$ (2.7) has a unique solution $p$ with the initial value $p(1) = 1$. In principle, the recursion (2.7) has probabilistic meaning for arbitrary $q$ since it determines a sequence of exchangeable partitions of $[n]$'s for $n$ in some finite or infinite range. Distributions of these partitions are obtained algebraically by fully expanding $p$ through $q$. However, typically, these partitions of $n$ are not consistent with respect to restrictions, so in the infinite case, they might not determine the distribution of a partition of $\mathbb{N}$.

**3. Coalescents with freeze.** We consider the structure of a partition of a set (resp., of an integer) with each of its blocks (or parts) assigned one of two possible conditions which we call *active* and *frozen*. We call such a combinatorial object a *partially frozen partition* of a set or of an integer. This added marking system makes it possible to provide a natural generalization of partition structures derived from a coalescent with Poisson mutations along the branches of a genealogical tree. We use the symbol $\Sigma_n^*$ for the pure singleton partition of $[n]$ with all blocks active and $\Sigma_\infty^*$ for the sequence $(\Sigma_n^*)_{n=1}^\infty$. As special cases of partially frozen partitions, we include the possibility that all blocks may be active, or all frozen. Ignoring the conditions of the blocks of a partially frozen partition, $\pi^*$ *induces* an ordinary partition $\pi$.

The $^*$-*shape* of a partially frozen partition $\pi_n^*$ of $[n]$ is the corresponding partially frozen partition of $n$ and the ordinary shape is defined in terms of the induced partition $\pi_n$.

For each positive integer $n$, we denote by $\mathcal{P}_{[n]}^*$ the set of all partially frozen partitions of $[n]$. Let $\mathcal{P}_\infty^*$ be the set of all partially frozen partitions of $\mathbb{N}$. We identify each element $\pi_\infty^* \in \mathcal{P}_\infty^*$ as the sequence $(\pi_1^*, \pi_2^*, \dots) \in \mathcal{P}_{[1]}^* \times \mathcal{P}_{[2]}^* \times \cdots$, where $\pi_n^*$ is $\pi_\infty^*|_n$, the restriction of $\pi_\infty^*$ to $[n]$. Endowing $\mathcal{P}_\infty^*$ with the topology it inherits as a subset of $\mathcal{P}_{[1]}^* \times \mathcal{P}_{[2]}^* \times \cdots$, the space $\mathcal{P}_\infty^*$ is compact and metrizable. We call a random partially frozen partition of $[n]$ *exchangeable* if its distribution is invariant under the action of permutations of $[n]$. Similarly to [9, 20], we call a $\mathcal{P}_\infty^*$-valued stochastic process $(\Pi_\infty^*(t), t \ge 0)$ a *coalescent* if it has càdlàg paths and $\Pi_\infty^*(s)$ is a $^*$-*refinement* of $\Pi_\infty^*(t)$ for every $s < t$, meaning that the induced partition $\Pi_\infty(s)$ is a refinement of $\Pi_\infty(t)$ and the set of frozen blocks of $\Pi^*_\infty(s)$ is a subset of the set of frozen blocks of $\Pi^*_\infty(t)$.

The construction of an exchangeable random partition of $\mathbb{N}$ by cutting branches of the merger-history tree of a $\Lambda$-coalescent $(\Pi_\infty(t), t \geq 0)$ by mutations with rate $\rho$ can now be formalized as follows. For each $i \in \mathbb{N}$, let $\tau_i$ denote the random time at which a mutation first occurs along the line of descent to leaf $i$ of the tree and declare the block of $\Pi_\infty(t)$ containing $i$ to be active if $\tau_i > t$ and frozen if $\tau_i \leq t$. This defines a $\mathscr{P}^*_\infty$-valued Markov process $(\Pi^*_\infty(t), t \geq 0)$. As $t \to \infty$, the state $\Pi^*_\infty(t)$ approaches a limit $\Pi^*_\infty(\infty)$ with all blocks frozen. This is the exchangeable random partition generated by the exchangeable sequence of random variables $(\tau_i, i \in \mathbb{N})$, meaning that two integers $i$ and $j$ are in the same block of $\Pi^*_\infty(\infty)$ if and only if $\tau_i = \tau_j$. Assuming that $\Pi^*_\infty(0) = \Sigma^*_\infty$, it should be clear that the EPPF of $\Pi^*_\infty(\infty)$ is that defined by Möhle's recursion (2.7). The following two theorems present more formal statements.

THEOREM 3.1. *Let $(\lambda_{b,k}, 2 \leq k \leq b < \infty)$, $(\rho_n, 1 \leq n < \infty)$ be two arrays of nonnegative real numbers. There exists, for each $\pi^*_\infty \in \mathscr{P}^*_\infty$, a $\mathscr{P}^*_\infty$-valued coalescent $(\Pi^*_\infty(t), t \geq 0)$ with $\Pi^*_\infty(0) = \pi^*_\infty$, whose restriction, for each $n$, $(\Pi^*_n(t), t \geq 0)$ to $[n]$ is a $\mathscr{P}^*_{[n]}$-valued Markov chain starting from $\pi^*_n = \pi^*_\infty|_n$, and which evolves with the rules*

- *at each time $t \geq 0$, conditionally given $\Pi^*_n(t)$ with $b$ active blocks, each $k$-tuple of active blocks of $\Pi^*_n(t)$ is merging to form a single active block at rate $\lambda_{b,k}$ and*
- *each active block turns into a frozen block at rate $\rho_{n,b}$*

*if and only if the integral representation (2.5) holds for some nonnegative finite measure $\Lambda$ on the Borel subsets of $[0, 1]$ and $\rho_{n,b} = \rho$ for some nonnegative real number $\rho$. This $\mathscr{P}^*_\infty$-valued process $(\Pi^*_\infty(t), t \geq 0)$ directed by $(\Lambda, \rho)$ is a strong Markov process. For $\rho = 0$, this process reduces to the $\Lambda$-coalescent and for $\rho > 0$, the process is obtained by superposing Poisson marks at rate $\rho$ on the merger-history tree of a $\Lambda$-coalescent and freezing the block containing $i$ at the time of the first mark along the line of descent of $i$ in the merger-history tree.*

PROOF. Just as in [29], consistency of the rate descriptions for different $n$ implies that (2.6) holds, hence the integral representation (2.5) and equality of the $\rho_{n,b}$'s are also obvious by consistency. $\square$

DEFINITION 3.2. Call this $\mathscr{P}^*_\infty$-valued Markov process directed by a nonnegative integer $\rho$ and a nonnegative finite measure $\Lambda$ on $[0, 1]$ the $\Lambda$-*coalescent freezing at rate* $\rho$, or the $(\Lambda, \rho)$-*coalescent* for short. Call a $(\Lambda, \rho)$-coalescent starting from state $\Sigma^*_\infty$ a *standard $\Lambda$-coalescent freezing at rate* $\rho$, where $\Sigma^*_\infty$ is the pure singleton partition with all blocks active.

Consider the finite coalescent with freeze $(\Pi_n^*(t), t \geq 0)$ which is the restriction of a standard $\Lambda$-coalescent freezing at rate $\rho$ to $[n]$. According to the description above, all active blocks will coalesce by the rules of a $\Lambda$-coalescent, except that every active block enters the frozen condition at rate $\rho$ and after that, the block will stay frozen forever. Hence, it is clear that as long as the freezing rate $\rho$ is positive, in finite time, the process $(\Pi_n^*(t), t \geq 0)$ will eventually reach a *final partition* $E_n^*$, with all of its blocks in the frozen condition.

Now, recall Möhle's model [23], as reviewed in Section 2. The ancestral lines of $n$ labeled genes of current generation coalesce as a $\Lambda$-coalescent and mutations occur along each ancestral line as Poisson point process with rate $\rho > 0$. Hence, the final partition of $[n]$ is defined so that if the ancestral line of an individual is interrupted by a mutation before the line coalesces with any other ancestral lines, the individual will be a singleton in the partition. This corresponds to the idea of freezing here: tracing the evolution of a particle starting from time 0, if a particle freezes before coalescing with others, it will enter as a singleton block in the final partition of the process.

Going into further detail, let us look at the discrete chain embedded in $\Lambda$-coalescent freezing at rate $\rho$. By definition, for each time $t \geq 0$, $\Pi_n^*(t)$ is a partially frozen exchangeable random partition of $[n]$, hence its induced form $\Pi_n(t)$ gives an exchangeable random partition of $[n]$, as does the final partition $E_n^* = \Pi_n^*(\infty)$ and its induced form $E_n$. Set $E_\infty^* := (E_n^*)$ as the final partition of $(\Pi_\infty^*(t), t \geq 0)$ and denote its induced partition by $E_\infty = (E_n)$.

Following the terminology in [15, 13], we call a triangular stochastic matrix a *decrement matrix*. We use the notation $q_n = (q(b:k); 1 \leq k \leq b \leq n)$ or $q_\infty = (q(n:k); 1 \leq k \leq n < \infty)$ to indicate whether we wish to consider finite or infinite matrices. Thus the entries of a decrement matrix are nonnegative and satisfy $\sum_{k=1}^{b} q(b:k) = 1$ for all $b$ in the required range. The following facts can be read from the existence of $(\Pi_\infty^*(t), t \geq 0)$ and Möhle's analysis recalled around (2.7).

THEOREM 3.3 ([23], Theorem 3.1). *The induced final partition $E_\infty = (E_n)_{n=1}^\infty$ of a standard $\Lambda$-coalescent freezing at rate $\rho > 0$ is an exchangeable infinite random partition of $\mathbb{N}$ whose EPPF $p$ is the unique solution of Möhle's recursion* (2.7) *with coefficients from the infinite decrement matrix $q_\infty$ defined through* $(\Lambda, \rho)$ *as in* (2.8).

**4. Freeze-and-merge operations.** Given a stochastic process $X$ indexed by a continuous-time parameter $t \geq 0$, assuming $X$ has right continuous piecewise constant paths, the *jumping process derived from* $X$ is the discrete-time process

$$\widehat{X} = (\widehat{X}(0), \widehat{X}(1), \ldots) = (X(T_0), X(T_1), X(T_2), \ldots),$$

where $T_0 := 0$ and $T_k$ for $k \geq 1$ is the least $t > T_{k-1}$ such that $X(t) \neq X(T_{k-1})$, if there is such a $t$, and $T_k = T_{k-1}$ otherwise. The processes $X$ of interest here will ultimately arrive in some absorbing state and then so too will $\widehat{X}$. In particular, the finite coalescent with freeze $(\Pi_n^*(t), t \geq 0)$, obtained by restriction to $[n]$ of a $\Lambda$-coalescent freezing at positive rate $\rho$, is a Markov chain with transition rate $\binom{b}{k}\lambda_{b,k}$ for a $k$-merge and rate $b\rho$ for a freeze, where $b$ is the number of active blocks at time $t$ and the $\lambda_{b,k}$'s are as in (2.5). The jumping process $\widehat{\Pi}_n^*$ is then a Markov chain governed by the following *freeze-and-merge* operation $\mathrm{FM}_n$ which acts on a generic partially frozen partition $\pi_n^*$ of $[n]$ as follows: if $\pi_n^*$ has $b > 1$ active blocks, then:

- with probability $q(b:k)$, some $k$ of $b$ active blocks are chosen uniformly at random and merged into a single active block (for $2 \leq k \leq b$);
- with probability $q(b:1)$, an active block is chosen uniformly at random from $b$ blocks and turned into a frozen block.

In the case $b = 1$, only the second option is possible, that is, $q(1:1) = 1$, and when all blocks of $\pi_n^*$ are in frozen condition, the operation is defined to be the identity. For the $\Lambda$-coalescent freezing at positive rate $\rho$, we know that

(i) the decrement matrix $q$ is of the special form (2.8);

(ii) the continuous-time processes $\Pi_n^*(t)$ are Markovian and consistent as $n$ varies, meaning that $\Pi_m^*(t)$ for $m < n$ coincides with $\Pi_n^*(t)|_m$, the restriction of $\Pi_n^*(t)$ to $[m]$.

Note that $\mathrm{FM}_n$ always reduces the number of active blocks. In particular, it transforms a partition of $[n]$ with $b > 1$ active blocks into some other partition of $[n]$ with $b - 1$ active blocks with probability $q(b:1) + q(b:2)$.

To view Möhle's recursion (2.7) in greater generality, we consider this freeze-and-merge operation $\mathrm{FM}_n$ for $n$ some fixed positive integer and $q_n$ a finite decrement matrix. Let $(\widehat{\Pi}_n^*(k), k = 0, 1, 2, \ldots)$ be the Markov chain obtained by iterating $\mathrm{FM}_n$ starting from $\widehat{\Pi}_n^*(0) = \Sigma_n^*$. Since $\mathrm{FM}_n$ is defined in terms of $^*$-shapes, each $\widehat{\Pi}_n^*(k)$ is a partially frozen exchangeable partition of $[n]$. The $\mathrm{FM}_n$-chain is strictly transient, in the sense that it never passes through the same state until it reaches a partially frozen partition $E_n^*$, all of whose blocks are frozen. Let $E_n$ be the induced partition of $[n]$, which we call the *final partition*, and consider $E_n$ as the outcome of a random transformation of exchangeable partitions $\Sigma_n \mapsto \Sigma_n^* \mapsto E_n^* \mapsto E_n$.

Observe that for $m = 1, \ldots, n$, the first $m$ rows of the decrement matrix $q_n$ comprise a decrement matrix $q_m$ which itself defines a freeze-and-merge operation $\mathrm{FM}_m$ on partially frozen partitions of $[m]$. Hence, for given $q_n$, we can also define a final partition $E_m$ of the $\mathrm{FM}_m$-chain. Note that $\mathrm{FM}_n$ is essentially an operation on the *set* of active blocks, regardless of their contents, sizes and the configuration of frozen blocks.

LEMMA 4.1. *Given an arbitrary decrement matrix $q_n$, let $p$ be the function on $\bigcup_{m=1}^n \mathcal{C}_m$ whose restriction to $\mathcal{C}_m$ is the EPPF of $E_m$, the final partition generated* *by the* $\mathrm{FM}_m$ *chain, for* $1 \le m \le n$. *Then* $p$ *satisfies Möhle's recursion* (2.7) *for each composition* $(n_1, n_2, \dots, n_\ell) \in \mathcal{C}_n$.

PROOF. A particular realization of $E_n$ with $\mathrm{shape}(E_n) = (n_1, \dots, n_\ell)$ occurs when either

(a) some block $\{j\}$ of $E_n$ appears as a frozen singleton in $\mathrm{FM}_n(\Sigma_n^*)$ and all other singletons $\{i\} \ne \{j\}$ evolve to form a partition with shape $(\dots, \widehat{1}, \dots)$, or
(b) the first iteration of $\mathrm{FM}_n$ merges some singletons $\{j_1\}, \dots, \{j_k\}$ $(k > 1)$ in a single active block which completely enters one of the blocks of $E_n$.

By the definition of $p$ and the last remark before the lemma, the probability of event (a) is

$$\frac{1}{n} \cdot q(n:1)\, p(\dots, \widehat{1}, \dots)$$

because after $\{j\}$ gets frozen, the operation $\mathrm{FM}_n$ is reduced to $\mathrm{FM}_{n-1}$ acting on partially frozen partitions of $[n] \setminus \{j\}$. Similarly, the probability of (b) is

$$\frac{1}{\binom{n}{k}} \cdot q(n:k)\, p(\dots, n-k+1, \dots)$$

because after creation of the active block $\{j_1, \dots, j_k\}$, the iterates of $\mathrm{FM}_n$ can be identified with those of $\mathrm{FM}_{n-k+1}$ acting on partially frozen partitions of $[n] \setminus \{j_2, \dots, j_k\}$. Summation over all possible choices yields (2.7). $\square$

In the general setting of Lemma 4.1, the sequence of exchangeable final partitions $(E_m)_{m=1}^n$ need not be consistent with respect to restrictions. We turn next to the constraints on $q$ imposed by the following, stronger, consistency condition.

DEFINITION 4.2. For a decrement matrix $q_n$ and $1 \le m < n$, call the transition operators $\mathrm{FM}_n$ and $\mathrm{FM}_m$ derived from $q_n$ *consistent* if whenever $\widehat{\Pi}_n^*$ is a Markov chain governed by $\mathrm{FM}_n$, the jump process derived from the restriction of $\widehat{\Pi}_n^*$ to $[m]$ is a Markov chain governed by $\mathrm{FM}_m$. Call the decrement matrix $q_n$ *consistent* if this condition holds for every $1 \le m < n$.

As the leading example, it is clear from consistency of the continuous-time chains $(\Pi_n^*(t),\ t \ge 0)$ which represent a $(\Lambda, \rho)$-coalescent that for every $n$, the corresponding decrement matrix $q_n$ is consistent. The following lemma collects some general facts about consistency. The proofs are elementary and left to the reader. Let $\mathrm{FM}_n(\pi_n^*)$ denote the random partition obtained by the action of $\mathrm{FM}_n$ on an initial partially frozen partition $\pi_n^*$ of $[n]$.

LEMMA 4.3. *Given a particular decrement matrix* $q_n$, *we have the following*:

(i) *For fixed* $1 \le m < n$, *the transition operators* $\mathrm{FM}_m$ *and* $\mathrm{FM}_n$ *are consistent if and only if for each partially frozen partition* $\pi_n^*$ *of* $[n]$, *there is the equality in distribution*

$$\mathrm{FM}_m(\pi_n^*|_m) \stackrel{d}{=} \mathrm{FM}_n(\pi_n^*)||_m,$$

*where on the left-hand side* $\pi_n^*|_m$ *is the restriction of* $\pi_n^*$ *to* $[m]$ *and on the right-hand side*, *the notation* $||_m$ *means the restriction to* $[m]$ *conditional on the event* $\mathrm{FM}_n(\pi_n^*|_m) \neq \pi_n^*|_m$ *that* $\mathrm{FM}_n$ *freezes or merges at least one of the blocks of* $\pi_n^*$ *containing some element of* $[m]$.

(ii) *If* $\mathrm{FM}_{m-1}$ *and* $\mathrm{FM}_m$ *are consistent for every* $1 < m \le n$, *then so are* $\mathrm{FM}_m$ *and* $\mathrm{FM}_n$ *for every* $1 < m \le n$; *that is*, $q_n$ *is consistent.*

LEMMA 4.4. *A decrement matrix* $q_n$ *is consistent if and only if it satisfies the backward recursion*

$$\begin{aligned}
q(b:k) &= \frac{k+1}{b+1}q(b+1:k+1) + \frac{b+1-k}{b+1}q(b+1:k)\\
&\quad + \frac{1}{b+1}q(b+1:1)q(b:k) + \frac{2}{b+1}q(b+1:2)q(b:k)\\
&\qquad\qquad (2 \le k \le b < n),
\end{aligned}\tag{4.1}$$

$$\begin{aligned}
q(b:1) &= \frac{b}{b+1}q(b+1:1) + \frac{1}{b+1}q(b+1:1)q(b:1)\\
&\quad + \frac{2}{b+1}q(b+1:2)q(b:1) \qquad (1 \le b < n).
\end{aligned}\tag{4.2}$$

*Consequently*, *each probability distribution* $q(n:\cdot)$ *on* $[n]$ *determines a unique consistent decrement matrix* $q_n$ *with this* $n$*th row.*

PROOF. Consider $\mathrm{FM}_n$ and $\mathrm{FM}_{n-1}$ applied to $\Sigma_n^*$ and $\Sigma_{n-1}^*$, that is, the partitions into singletons, all in the active condition. For $k \le n-1$, $\mathrm{FM}_{n-1}$ operates by coalescing $\{1, \dots, k\}$ into an active block with probability

$$\frac{q(n-1:k)}{\binom{n-1}{k}}.\tag{4.3}$$

As for the jumping process of ($\mathrm{FM}_n$ restricted to $[n-1]$), the probability of a coalescence of $\{1, \dots, k\}$ into an active block is the sum of the following four parts, depending on the development of the $\mathrm{FM}_n$ chain. Let $T_1$ be the time of the first change in the restriction of the $\mathrm{FM}_n$ chain to $[n-1]$. To obtain the required coalescence, either (i) $T_1 = 1$ and the state after a single step of $\mathrm{FM}_n$ comes from $\Sigma_n^*$ by coalescing $\{1, \dots, k, n\}$ or $\{1, \dots, k\}$, these occurring with probability

$$\frac{q(n:k+1)}{\binom{n}{k+1}} + \frac{q(n:k)}{\binom{n}{k}},\tag{4.4}$$

or (ii) $T_1 = 2$ with $\mathrm{FM}_n$ acting on $\Sigma_n^*$ by first freezing $\{n\}$ then coalescing $\{1, 2, \ldots, k\}$, or first coalescing $\{n\}$ with one of the other $n-1$ singletons, leaving $1, 2, \ldots, k$ in $k$ distinct blocks, then coalescing these $k$ blocks at the next step, these occurring with probability

$$\frac{q(n:1)}{n} \cdot \frac{q(n-1:k)}{\binom{n-1}{k}} + \frac{(n-1)q(n:2)}{\binom{n}{2}} \cdot \frac{q(n-1:k)}{\binom{n-1}{k}}. \tag{4.5}$$

Equate (4.3) with the sum of (4.4) and (4.5) to get (4.1) for $b = n-1$. In much the same way, $\mathrm{FM}_{n-1}$ may act on $\Sigma_{n-1}^*$ by freezing $\{1\}$ with probability

$$\frac{q(n-1:1)}{n-1}. \tag{4.6}$$

While for the jumping process of ($\mathrm{FM}_n$ restricted to $[n-1]$), to obtain the required form, either (i) $T_1 = 1$ and $\mathrm{FM}_n$ acts on $\Sigma_n^*$ by freezing $\{1\}$ with probability

$$\frac{q(n:1)}{n}, \tag{4.7}$$

or (ii) $T_1 = 2$ and the result is obtained from $\Sigma_n^*$ by first freezing $\{n\}$ then freezing $\{1\}$, or first coalescing $\{n\}$ with one of other $n-1$ singletons then freezing the block containing 1, these ways occurring with probability

$$\frac{q(n:1)}{n} \cdot \frac{q(n-1:1)}{n-1} + \frac{(n-1)q(n:2)}{\binom{n}{2}} \cdot \frac{q(n-1:1)}{n-1}. \tag{4.8}$$

Equate (4.6) with the sum of (4.7) and (4.8) to get (4.1) for $b = n-1$. Combine them to get (4.2) for $b = n-1$. The recursions for $b < n$ follow by replacing $n$ by $b+1$.

Conversely, granted the recursions (4.1) and (4.2), in order to prove consistency, it is sufficient to check the case $m = n-1$ and this is done by application of Lemma 4.3. □

LEMMA 4.5. *For $1 \le m \le n$, let $E_m$ be the final partition of the* $\mathrm{FM}_n$*-chain starting in state $\Sigma_m^*$. If the decrement matrix $q_n$ is consistent, then the finite sequence of exchangeable random set partitions $(E_m)_{m=1}^n$ is consistent in the sense that*

$$E_m \stackrel{d}{=} E_n|_m.$$

*The finite EPPF $p$ of $(E_m)_{m=1}^n$ then satisfies Möhle's recursion* (2.7) *for all compositions of $m \le n$ in the left-hand side.*

PROOF. The consistency in distribution is clear. To show (2.7), it suffices to look at the case with compositions of $n$ on the left-hand side, for which Lemma 4.1 applies. □

Here is our principal result regarding finite partitions which satisfy (2.7).

THEOREM 4.6. *For a positive integer $n > 1$ and arbitrary probability distribution $q(n:\cdot)$ on $[n]$,*

(i) *there exists a unique finite EPPF $p$ for a consistent sequence of random set partitions $(\Pi_m)_{m=1}^n$ which satisfies Möhle's recursion* (2.7) *for all compositions of $n$;*

(ii) *this finite EPPF $p$ satisfies Möhle's recursion* (2.7) *for all compositions of positive integers $m < n$ with coefficients $q(m:\cdot)$ derived from $q(n:\cdot)$ by the recursion* (4.1), (4.2);

(iii) *for each $1 \leq m \leq n$, the distribution of $\Pi_m$ determined by the restriction of this EPPF $p$ to compositions of $m$ is that of the final partition of the* $\mathrm{FM}_m$ *Markov chain with decrement matrix $q_m$ defined by* (ii), *starting from state $\Sigma_m^*$.*

PROOF. We apply Lemma 4.5. Given arbitrary probability distribution $q(n:\cdot)$ on $[n]$, we can define all $q(m:\cdot)$, $1 \leq m < n$, by the backward recursion (4.1) and (4.2). We then use the decrement matrix $q_n$ with these rows to build a sequence of Markov chains: for each $m$, the chain $(\Pi_m(k), k = 0, 1, 2, \dots)$ starts from $\Sigma_m^*$ and evolves according to $\mathrm{FM}_m$. The sequence of induced final partitions $(E_m)_{m=1}^n$ of these chains has EPPF $p$ which satisfies recursion (2.7). Hence, the existence part of (i) follows. We postpone the proof of uniqueness in part (i) to the next section. The assertions (ii) and (iii) follow directly from this construction. $\square$

**5. The sample-and-add operation.** Given a probability distribution $q(n:\cdot)$ on $[n]$, we now interpret Möhle's recursion (2.7) as the system of equations for the invariant probability measure of a particular Markov transition mechanism on partitions of $[n]$ and show that this invariant probability distribution is unique. This will complete the proof of Theorem 4.6.

Consider the following *sample-and-add* random operation on $\mathcal{P}_{[n]}$, denoted by $\mathrm{SA}_n$. We regard a generic random partition $\Pi_n \vdash [n]$ as a random allocation of balls labeled $1, \dots, n$ to some set of nonempty boxes, which the operation $\mathrm{SA}_n$ transforms into some other random allocation $\Pi_n'$. Fix $q(n:\cdot)$, a probability distribution on $[n]$, and let $K_n$ be a random variable with this distribution $q(n:\cdot)$. Given $K_n = k$ and $\Pi_n = \pi_n$, we have the following:

- if $k = 1$, first delete a single ball picked uniformly at random from the balls allocated according to $\pi_n$ to make an intermediate partition of some set of $n - 1$ balls, then add to this intermediate partition a single box containing the deleted ball;
- if $k = 2, \dots, n$, delete a sequence of $k - 1$ of the $n$ balls from $\pi_n$ by uniform random sampling without replacement to obtain an intermediate partition of some set of $n - k + 1$ balls, then mark a ball picked uniformly from these $n - k + 1$ balls and add the $k - 1$ sampled balls to the box containing the marked ball.

In either case, empty boxes are deleted in case any appear after the sampling step. The resulting partition of $[n]$ is $\Pi'_n$. For each $q(n:\cdot)$, this defines a Markovian transition operator $\mathrm{SA}_n$ on partitions of $[n]$.

LEMMA 5.1. *Let $\Pi_n$ be an exchangeable random partition of $[n]$ with finite EPPF $p$ defined as a function of compositions of $m$ for $1 \le m \le n$. Let $\Pi'_n$ be derived from $\Pi_n$ by the $\mathrm{SA}_n$ operation determined by some arbitrary probability distribution $q(n:\cdot)$ on $[n]$. Then $\Pi'_n$ is an exchangeable random partition of $[n]$ whose EPPF $p'$ is determined on compositions of $[n]$ by the formula*

$$
\begin{aligned}
p'(n_1, n_2, \ldots, n_\ell) = {} & \frac{q(n:1)}{n} \sum_{j:n_j=1} p(\ldots, \widehat{n_j}, \ldots) \\
& + \sum_{k=2}^{n} q(n:k) \sum_{j:n_j \ge k} \frac{\binom{n_j}{k}}{\binom{n}{k}} p(\ldots, n_j - k + 1, \ldots).
\end{aligned} \tag{5.1}
$$

[Note that the right-hand side of (5.1) is identical to the right-hand side of Möhle's recursion (2.7).]

PROOF. Let $K_n$ with distribution $q(n:\cdot)$ be the number of balls deleted in the $\mathrm{SA}_n$ operation. For each partition $\pi'_n$ of $[n]$, we can compute

$$
\mathbb{P}(\Pi'_n = \pi'_n) = \sum_{k=1}^{n} q(n:k) \mathbb{P}(\Pi'_n = \pi'_n | K_n = k). \tag{5.2}
$$

Assuming that $\pi'_n$ has boxes of size $n_1, \ldots, n_\ell$ and that the $\mathrm{SA}_n$ operation acts on an exchangeable $\Pi_n$ with EPPF $p$, we deduce (5.1) from (5.2) and

$$
\mathbb{P}(\Pi'_n = \pi'_n | K_n = 1) = \frac{1}{n} \sum_{j:n_j=1} p(\ldots, \widehat{n_j}, \ldots), \tag{5.3}
$$

$$
\mathbb{P}(\Pi'_n = \pi'_n | K_n = k) = \sum_{j:n_j \ge k} \frac{\binom{n_j}{k}}{\binom{n}{k}} p(\ldots, n_j - k + 1, \ldots), \qquad k \ge 2. \tag{5.4}
$$

First, consider (5.4). For the event $(\Pi'_n = \pi'_n)$ to occur, there must be some $j$ with $n_j \ge k$. For each such $j$, corresponding to a box of $\pi'_n$ with at least $k$ balls, the result $(\Pi'_n = \pi'_n)$ might be obtained by the addition of $k-1$ balls to that box. The sequence of labels of these balls, in order of their choice, can be any one of $n_j(n_j-1)\cdots(n_j-k+2)$ sequences and the final ball chosen to mark the box can be any one of $n_j - k + 1$ balls, making $k!\binom{n_j}{k}$ choices out of a total of $k!\binom{n}{k}$ possible choices. Given one of these $k!\binom{n_j}{k}$ choices of $k$ balls, let $M_{k-1}$ be the set of labels of the $k-1$ balls that are moved. Then the event $(\Pi'_n = \pi'_n)$ occurs if and only if the restriction of $\Pi_n$ to $[n] - M_{k-1}$ equals the restriction of $\pi'_n$ to $[n] - M_{k-1}$, which is a particular partition of $n-k+1$ labeled balls into boxes of $\bar{n}_1, \dots, \bar{n}_\ell$ balls, where $\bar{n}_i = n_i 1(i \neq j) + (n_j - k + 1)1(i = j)$. The conditional probability of $(\Pi'_n = \pi'_n)$, given $K_n = k$ and which of the $k!\binom{n_j}{k}$ possible choices of $k$ balls is made, is therefore $p(\dots, n_j - k + 1, \dots)$, by the assumed exchangeability of $\Pi_n$ and the definition of the EPPF $p$ of $\Pi_n$ on compositions of $m \le n$ by restriction of $\Pi_n$ to subsets of size $m$. The evaluation (5.4) is now apparent and (5.3) is also apparent by a similar, but easier, argument. □

PROPOSITION 5.2. *For each probability distribution* $q(n:\cdot)$ *on* $[n]$, *the corresponding* $\mathrm{SA}_n$ *transition operator on partitions of* $[n]$ *has a unique stationary distribution. A random partition with this stationary distribution is exchangeable and its EPPF is the finite unique EPPF* $p$ *that satisfies Möhle's recursion* (2.7), *that is*, (5.1) *with* $p' = p$.

PROOF. If $q(n:1) = 1$, then $\mathrm{SA}_n$ eventually terminates with singleton partition, so the stationary distribution is degenerate and concentrated on the singleton partition. If $q(n:1) = 0$, then $\mathrm{SA}_n$ eventually terminates with one-block partition, so the stationary distribution is degenerate and concentrated on the one-block partition. If $0 < q(n:1) < 1$, then we also have $q(n:k) > 0$ for some $k > 1$; in this case, the stationary law is again unique because all states communicate (e.g., the pure-singleton partition $\Sigma_n$ is reachable from everywhere and can reach any partition in finitely many steps, as is easily verified). Observe that passing to shapes projects the $\mathrm{SA}_n$ chain with state space partitions of the set $[n]$ onto another Markov chain whose state space is the set of partitions of the integer $n$. It easily follows that the unique stationary distribution of $\mathrm{SA}_n$ governs an exchangeable random partition of $[n]$. The previous lemma shows that its EPPF $p$ solves Möhle's recursion. Finally, if an EPPF $p$ solves Möhle's recursion, then it provides a stationary state for the $\mathrm{SA}_n$ chain. Hence, the uniqueness result follows for solutions of Möhle's recursion by an EPPF $p$. □

5.1. *Special cases.* The following are two special cases of $\mathrm{SA}_m$ operation.

*Ewens' partition* [1, 10] appears when $q(n:\cdot)$ may have only two positive entries,

$$q(n:1) = \frac{2\rho}{n-1+2\rho} \quad \text{and} \quad q(n:2) = \frac{n-1}{n-1+2\rho}$$

for each $n \ge 2$. It is easy to realize that the $\mathrm{SA}_n$ operation in this case is reduced to the following operation with $u = 2\rho/(n-1+2\rho)$: given a number $0 \le u \le 1$ and a partition of $[n]$ as allocation of $n$ labeled balls, we first uniformly sample two balls named $A$ and $D$ without replacement from the $n$ balls (so $A = D$ is excluded), then put ball $A$ back where it was, and finally:

- with probability $u$, append a new box containing the single ball $D$,

- with probability $1-u$, add the ball $D$ to the box containing the ball $A$.

In this case, if we consider the FM operator determined by $q$, it is clear that only binary merges happen. That the stationary partition $\Pi_n$ follows the Ewens' sampling formula with parameter $\theta=(n-1)u/(1-u)$ is seen by the "Chinese restaurant" rule [28] for transition from $\Pi_{n-1}$ to $\Pi_n$, or can easily be concluded from the formula. The coincidence of the stationary distribution of this $\mathrm{SA}_n$ chain with the law of the induced final partition $E_n$ of the associated $\mathrm{FM}_n$ chain confirms, in this case, the well-known fact that Kingman's coalescent with mutations terminates at Ewens' partition.

The $\mathrm{SA}_n$-chain resembles Moran's novel mutation chain [26, 35, 37]. Transitions of the latter are the following: given a number $0\le u\le 1$ and a partition of $[n]$ as allocation of labeled balls, first choose two balls named $A$ and $D$ uniformly and independently from the $n$ balls (so $A=D$ is not excluded), then follow the rules

- with probability $u$, append a new box with a single ball $C$,
- with probability $1-u$, add a ball $C$ to the box that contains ball $A$,

then assign to ball $C$ the same label as that of $D$ and finally remove ball $D$. It is well known [35] that the stationary law of Moran's chain corresponds to Ewens' partition with parameter $nu/(1-u)$.

*Hook partitions.* Another extreme case appears when $q(n:\cdot)$ may have only two positive entries,

$$q(n:1)=\frac{n\rho}{1+n\rho}\quad\text{and}\quad q(n:n)=\frac{1}{1+n\rho}.$$

In this case, $\mathrm{SA}_n$ creates some number of singletons and then after some number of steps puts all balls in a single box. If $0<q(n:1)<1$, the stationary distribution concentrates on partitions with a hook shape $(m,1,1,\dots,1)$. This partition results from the $\Lambda$-coalescent with freeze when $\Lambda=\delta_1$ is a Dirac mass at 1.

**6. Infinite partitions.** In this section, we pass from finite partitions to the projective limit and arrive at the desired integral representation of infinite decrement matrix $q_\infty$ satisfying recursion (4.1) and (4.2). This gives another approach to Möhle's partitions via consistent freeze-and-merge chains, which may be seen as discrete-time jumping processes associated with the $\Lambda$-coalescent with freeze.

An infinite sequence of freeze-and-merge operations $\mathrm{FM}:=(\mathrm{FM}_n, n=1,2,\dots)$ which satisfies the condition in Definition 4.2 for all positive integers $1\le m<n<\infty$ is called *consistent*. By Lemma 4.4, such a sequence FM is determined by an infinite decrement matrix $q_\infty$ which satisfies the recursion (4.1), (4.2).

For each $n=1,2,\dots,$ the Markov chain starting from $\Sigma_n^*$ and driven by $\mathrm{FM}_n$ terminates with an induced final partition $\Pi_n$. These comprise an infinite partition $\Pi_\infty=(\Pi_n)_{n=1}^\infty$ which we call the *final partition* associated with consistent FM. In the case $q(2:1)=0$, the final partition is the trivial one-block partition.

LEMMA 6.1. *For every infinite decrement matrix $q_\infty$ with entries satisfying the recursion* (4.1), (4.2), *there exist a nonnegative finite measure $\Lambda$ on* $[0, 1]$ *and a nonnegative real number $\rho$ which satisfy $(\Lambda, \rho) \neq (0, 0)$ and are such that the representation $q(n:k) = \Phi(n:k)/\Phi(n)$ $(1 \leq k \leq n)$ holds with $\Phi$ as in* (2.9)–(2.11). *The data $(\Lambda, \rho)$ are unique up to a positive factor.*

PROOF. Suppose that $q$ solves (4.1), (4.2) and that $q(2:2) < 1$. Let $\Phi(n)$, $n = 1, 2, \ldots,$ satisfy

$$\Phi(n)/\Phi(n+1) \text{ written as } \frac{\Phi(n)}{\Phi(n+1)} = 1 - \frac{1}{n+1} q(n+1:1) - \frac{2}{n+1} q(n+1:2) \tag{6.1}$$

for $n \geq 1$; because the right-hand side is strictly positive, this recursion has a unique solution with some given initial value $\Phi(1) = \rho$, where $\rho > 0$. For $2 \leq k \leq n$, set

$$\Phi(n:k) := q(n:k)\Phi(n),$$

then from (6.1) and (4.1),

$$\Phi(n:k) = \frac{k+1}{n+1}\Phi(n+1:k+1) + \frac{n+1-k}{n+1}\Phi(n+1:k) \qquad (2 \leq k \leq n < \infty).$$

Apart from a shift by 2, this is the well-known Pascal-triangle recursion appearing in connection with de Finetti's theorem and the Hausdorff moment problem, hence (2.10) holds for some nonnegative measure $\Lambda$ on Borel sets of $[0, 1]$. From (4.2), we find that

$$\rho = \frac{\Phi(1)q(1:1)}{1} = \cdots = \frac{\Phi(n)q(n:1)}{n} = \cdots$$

and from

$$\sum_{k=1}^{n} \Phi(n)q(n:k) = \Phi(n),$$

we deduce (2.11) and $q(n:1) = \rho n/\Phi(n)$. Setting, by definition, $\Phi(n:1) := \rho n$, we are done. For the special case $q(2:2) = 1$, it is easy to observe that $\rho = 0$ and we get $\Lambda = \delta_0$ by means of similar analysis. □

Recording this lemma together with previous results, we have the following.

THEOREM 6.2. *Let $(\Pi_n)_{n=1}^{\infty}$ be a nontrivial exchangeable random partition of $\mathbb{N}$, different from the trivial one-block partition. The following are then equivalent*:

(i) *the EPPF $p$ satisfies the recursion* (2.7) *with some infinite decrement matrix $q_\infty$*;

(ii) *this matrix is representable as* $q(n:k) = \Phi(n:k)/\Phi(n)$ *with* $\Phi$ *defined by* (2.9)–(2.11) *and some nontrivial* $(\Lambda, \rho)$, *which is unique up to a positive factor*;

(ii) *this* $\Pi_\infty$ *is induced by the final partition of some standard* $\Lambda$*-coalescent freezing at rate* $\rho$;

(iii) *this* $\Pi_\infty$ *is the final partition of some consistent* FM *operation*.

Complementing this result, we have the following uniqueness assertion.

LEMMA 6.3. *The correspondence* $q \mapsto p$ *between infinite decrement matrices with* $q(2:1) > 0$ *satisfying consistency* (4.1), (4.2) *and the EPPF's is bijective*.

PROOF. We only need to show that $p$, which by Lemma 4.5 must solve (2.7), uniquely determines $q$. For general infinite partitions, $q(2:1) = p(1,1) > 0$ implies that $p(1,1,\dots,1) > 0$. This applied to the singleton shapes, together with

$$p(1,\dots,1) = q(n:1)q(n-1:1)\cdots q(2:1),$$

shows that the $q(n:1)$'s are uniquely determined by $p$. To show that $q(n:m)$ for $1 \le m \le n-1$ is also determined by $p$, we exploit the formula

$$\begin{aligned} p(m,1,\dots,1) = \frac{q(n:m)}{\binom{n}{m}} p(1,\dots,1) + \sum_{k=2}^{m-1} q(n:k)\frac{\binom{m}{k}}{\binom{n}{k}} p(m-k+1,1,\dots,1) \\ + q(n:1)\frac{n-m}{n} p(m,\widehat{1},1,\dots,1) \end{aligned}$$

and argue by induction in $m = 2, 3, \dots, n-1$. $\square$

Thus if an exchangeable infinite partition can be realized as the induced final partition of a consistent FM-operation, then this FM-operation is unique. The realization via a $(\Lambda, \rho)$-coalescent process is unique up to a positive multiple of the parameters, which corresponds to a linear time-change of the coalescent. If there is no freeze, then the uniqueness fails since any $\Lambda$-coalescent terminates with the trivial one-block partition.

We classify the cases where some of the entries of $q$ are zeros. It is assumed that the starting partition is $\Sigma_\infty^*$.

(i) If $q(n:1) = 1$ holds for $n = 2$, then the same holds for $n \ge 2$. This is the pure-freeze coalescent with $\Lambda = 0$, hence $E_\infty = \Sigma_\infty$.

(ii) If $q(n:1) = 0$ holds for $n = 2$, then the same holds for $n \ge 2$. This is a $\Lambda$-coalescent with no freeze, hence $E_\infty$ is the one-block partition.

(iii) If $q(n:1) > 0$, $q(n:2) > 0$ and $q(n:1) + q(n:2) = 1$ hold for $n = 3$, then the same relations hold for $n \ge 3$. This is the case of Kingman's coalescent with freeze, where $\Lambda$ is a positive mass at 0 and $E_\infty$ is Ewens' partition.

(iv) If $q(n:1) > 0$, $q(n:n) > 0$ and $q(n:1) + q(n:n) = 1$ hold for $n = 3$, then they also hold for $n \geq 3$. In this case, $\Lambda$ is a positive mass at 1 and $E_\infty$ is a hook partition.

The "generic" case is characterized by $q(3:1) > 0$, $q(3:2) > 0$, $q(3:3) > 0$, in which case $0 < q(n:m) < 1$ for all $1 \leq m \leq n < \infty$.

**7. Positivity.** This section provides a construction of decrement matrices $q_\infty$ satisfying the consistency condition (4.1), (4.2), from a single sequence of real numbers satisfying a positivity condition. For $(c(n), n = 0, 1, 2, \ldots)$ a sequence of real numbers, the backward difference operator $\nabla$ is defined as

$$\nabla c(n) := c(n) - c(n+1)$$

and for any $j = 0, 1, 2, \ldots$, its iterates act as

$$\nabla^j c(n) = \sum_{i=0}^{j} (-1)^i \binom{j}{i} c(n+i).$$

Now, let $(\Phi(n), n = 1, 2, \ldots)$ be a sequence of real numbers and let $\rho$ be a positive real number. Define, for each $n$,

$$\Phi(n:1) := \rho n \tag{7.1}$$

and

$$\overline{\Phi}(n) := \Phi(n) - \rho n. \tag{7.2}$$

Define

$$\Psi(n) := \frac{\nabla \overline{\Phi}(n)}{n} \tag{7.3}$$

and let

$$\Phi(n:m) := -\binom{n}{m} \nabla^{m-2} \Psi(n-m+1), \qquad 2 \leq m \leq n. \tag{7.4}$$

With these definitions, it can be verified that for each $n$,

$$\Phi(n) = \Phi(n:1) + \Phi(n:2) + \cdots + \Phi(n:n). \tag{7.5}$$

Hence, if all $\Phi(n)$ are positive and all $\Phi(n:m)$ are nonnegative, then the matrix with entries

$$q(n:m) := \frac{\Phi(n:m)}{\Phi(n)}, \qquad 1 \leq m \leq n, \tag{7.6}$$

is a well-defined infinite decrement matrix. More than that, we have the following observation.

LEMMA 7.1. *Suppose that a sequence of positive real numbers* $\rho$, $\Phi(n)$, $n = 1, 2, \dots,$ *is such that each entry* $\Phi(n:1)$, $\Phi(n:m)$ *in* (7.1), (7.4) *is nonnegative. Then the matrix* (7.6) *satisfies the recursion* (4.1), (4.2).

PROOF. The definition (7.4) of $\Phi(n:m)$ implies the recursion

$$\Phi(n:m) = \frac{m+1}{n+1}\Phi(n+1:m+1) + \frac{n-m+1}{n+1}\Phi(n+1:m), \qquad 2 \le m \le n. \tag{7.7}$$

Using this relation, the first recursion (4.1) can be reduced to

$$2\Phi(n+1:2) = (n+1)\big(\Phi(n+1) - \Phi(n)\big) - \Phi(n+1:1),$$

which follows from the definition of $\Phi(n+1:2)$ and $\Phi(n+1:1)$. The second recursion is actually the definition of $\Phi(n+1:2)$ after we substitute in all of the $\Phi(n:1)$, $\Phi(n+1:1)$ terms. $\square$

The above lemma shows that given a sequence of positive real numbers with some additional positivity property, we can recover Möhle's partition structure by first defining a consistent decrement matrix and then using the recursion (2.7). By Lemma 6.1, we know that every decrement matrix satisfying consistency condition (4.1), (4.2) has an integral representation which is unique up to a positive factor, so it is clear that we also have integral representation for the sequence of $\Phi(n)$ given here.

PROPOSITION 7.2. *A sequence of positive real numbers* $\rho$, $\Phi(n)$, $n = 1, 2, \dots,$ *is such that each entry* $\Phi(n:1)$, $\Phi(n:m)$ *as in* (7.1)–(7.4) *is nonnegative if and only if these numbers admit the integral representation* (2.9)–(2.11) *for some nonnegative finite measure* $\Lambda$ *on* $[0, 1]$, *which is then unique*.

**8. Freezing times.** In this section, $(\Pi^*(t), t \ge 0)$ is a standard $(\Lambda, \rho)$-coalescent, with $(\Pi(t), t \ge 0)$-induced ordinary partition-valued process and final partition $E_\infty$. We assume that both $\Lambda$ and $\rho$ are nonzero. The process $(\Pi^0(t), t \ge 0)$ will denote the standard $\Lambda$-coalescent. We presume that all $(\Lambda, \rho)$-coalescents are defined consistently as $\rho$ varies so that the $\Pi(t)$'s and $E_\infty$ get finer as the freezing rate $\rho$ increases, in particular, each partition $\Pi(t)$ being finer than $\Pi^0(t)$ for each $t \ge 0$ and $\rho > 0$.

8.1. *Age ordering.* Assigning each individual $j \in \mathbb{N}$ the *freezing time* $\tau_j$ (when the active block containing $j$ gets frozen), the final partition $E_\infty$ is defined by sending $i, j$ to the same block if and only if $\tau_i = \tau_j$. The correspondence $j \mapsto \tau_j$ induces a total order on the set of blocks of $E_\infty$: we say that the block containing $j$ is *older* than the block containing $i$ if $\tau_i < \tau_j$. With this *age ordering*, $E_\infty$ is an *ordered exchangeable partition* of $\mathbb{N}$, as studied in [7, 8, 12, 14, 15].

We preserve the notation $E_\infty = (E_n)$ to denote the partition with this additional feature of total order on the set of blocks. The law of ordered partition $E_\infty$ is determined by an exchangeable composition probability function (ECPF) $c(n_1, \dots, n_\ell)$ on compositions of $n$. The ECPF $c$ must satisfy an addition rule similar to (2.4), but, unlike $p$, need not be symmetric. The EPPF $p$ of unordered partition is recovered from $c$ by symmetrization; see [15] for details.

With each $j$, we associate a random open interval $]a_j, b_j[$, where

$$
\begin{aligned}
a_j &= \lim_{n\to\infty} \#\{i \le n : \tau_i < \tau_j\}/n, \\
b_j - a_j &= \lim_{n\to\infty} \#\{i \le n : \tau_i = \tau_j\}/n
\end{aligned}
\tag{8.1}
$$

and the existence of the frequencies is guaranteed by de Finetti's theorem. Thus $a_j$ is the total frequency of blocks preceding the block containing $j$ and $b_j - a_j$ is the frequency of the block containing $j$. The random open set $U = \bigcup_j ]a_j, b_j[$ is the *paintbox* representing $E_\infty$. The partition $E_\infty$ can be uniquely recovered from $U$ by a simple sampling scheme [12, 15, 19, 20].

For instance, when $\Lambda = \delta_0$, the complement closed set is

$$U^c = \{1, Y_1, Y_1 Y_2, \dots, 0\}$$

for $Y_k$'s independent random variables whose distribution is beta$(2\rho, 1)$. This case has been thoroughly studied [7, 8] and it is well known that the arrangement of the block sizes in age order is inverse to the arrangement in size-biased order. In the case $\Lambda = \delta_1$, the set $U$ has only one interval, $]Y, 1[$, where $Y$ has a beta distribution.

8.2. *Properties of the final partition.* Some properties of $U$ for a $(\Lambda, \rho)$-coalescent with $\rho > 0$ follow from known results about the $\Lambda$-coalescents [29]. We shall only discuss the case $\Lambda\{1\} = 0$ since the case $\Lambda\{1\} > 0$ only differs by an independent exponential killing and its properties easily follow those that in the case $\Lambda\{1\} = 0$. Let

$$\mu_r := \int_0^1 x^r \Lambda(dx).$$

Denote by Leb the Lebesgue measure on $[0, 1]$. In the event $\mathrm{Leb}(U) < 1$, the ordered partition $E_\infty$ with paintbox $U$ has a positive total frequency of singletons blocks and in the event $\mathrm{Leb}(U) = 1$, there are no singleton blocks at all.

PROPOSITION 8.1. *If $\mu_{-1} < \infty$, then with probability one*:

(i) $\Pi^0(t)$ *has singletons for each* $t > 0$;
(ii) $\Pi^*(t)$ *has active singletons for each* $t > 0$;
(iii) $\Pi^*(t)$ *has frozen singletons for each* $t > 0$;

(iv) $E_\infty$ *has singleton blocks*.

*If* $\mu_{-1} = \infty$, *then the opposites of* (i)–(iv) *hold with probability one*.

PROOF. By [29], Lemma 25, if $\mu_{-1} < \infty$, then $\Pi^0(t)$ has singletons almost surely and if $\mu_{-1} = \infty$, the partition has no singletons almost surely. Now, if $\Pi^0(t)$ has singletons, each of them is active with probability $0 < e^{-\rho t} < 1$, independently of the others, thus the partially frozen partition $\Pi^*(t)$ has singletons in both conditions and the frozen ones are also singleton blocks of $E_\infty$. Conversely, if, with positive probability, $E_\infty$ has singletons, then for some $t$ with positive probability, $\Pi^*(t)$ has frozen singletons, then, perhaps for some other $t$, with positive probability, $\Pi^*(t)$ has active singletons, but in this event, the partition $\Pi^0(t)$ has singletons, hence $\mu_{-1} = \infty$ cannot hold. □

By [29], Proposition 23, the $\Lambda$-coalescent either *comes down from infinity* [the number of blocks in $\Pi^0(t)$ is finite almost surely for every $t > 0$] or *stays infinite* (the number of blocks is finite).

PROPOSITION 8.2. *If the* $\Lambda$-*coalescent stays infinite*, *then the* $(\Lambda, \rho)$-*coalescent has infinitely many active blocks at any time*, *therefore*:

(i) *the set of freezing times* $\{\tau_j\}$ *is dense in* $\mathbb{R}_+$;
(ii) *the closed set* $U^c$ *has empty interior and no isolated points*.

*If the* $\Lambda$-*coalescent comes down from infinity*, *then the* $(\Lambda, \rho)$-*coalescent satisfies*:

(i′) *the set of freezing times* $\{\tau_j\}$ *is bounded and only accumulates near* 0;
(ii′) *the closed set* $U^c$ *only accumulates near* 0.

PROOF. Let $J_k$ be the minimal element in some block $A_k$ of $\Pi^0(t)$. Then $J_k$ is also the minimal element in some block $B_k \subset A_k$ of $\Pi^*(t)$. Since the block containing $J_k$ changes the condition from active to frozen independently of the $\Lambda$-coalescent, with positive probability $1 - e^{-\rho t}$, the block $B_k$ is active. For $k = 1, 2, \dots$, these events are independent, hence $\Pi^*(t)$ has infinitely many active blocks. But the same is true for $t + \varepsilon$, hence, arguing as in Proposition 8.1, we see that infinitely many of the active $B_k$'s get frozen before $t + \varepsilon$, hence (i). Moreover, infinitely many of the active $B_k$'s are nonsingleton and hence, by the law of large numbers for exchangeable trials, have positive frequency. Assertion (ii) now follows from this remark, (i) and (8.1). □

**9. Comparison with regenerative partitions and Markovian fragmentations.** This section is mainly devoted to parallels and differences between Möhle's partitions and regenerative partitions [13–15]. A novel feature discussed here is a realization of regenerative partitions by a simple continuous-time coalescent process.

9.1. *Continuous-time realization and EPPF.* Consider a $\mathscr{P}^*_\infty$-valued Markovian process $(\Pi^*_\infty(t), t \geq 0)$ which starts with $\Sigma^*_\infty$ and evolves by the following rules. Any number of *active singleton* blocks can merge to form a *single frozen* block, which suspends further evolution immediately. In particular, an active singleton block can turn into a frozen singleton block, an event interpreted as unary merge. If $\Pi_n(t)$ has $b$ active blocks, then each $k$-tuple is merging at the same rate so that the total rate for a $k$-merge is $\Phi(b:k)$ for $1 \leq k \leq b < \infty$, and $\Phi(1:1) > 0$.

Eventually, there are only frozen blocks whose configuration determines a final partition $E_\infty$. Setting $\Phi(b) := \Phi(b:1) + \cdots + \Phi(b:b)$ and $q(n:k) := \Phi(n:k)/\Phi(n)$, the EPPF of $E_\infty$ satisfies

$$p(n_1, n_2, \dots, n_\ell) = \sum_{j=1}^{\ell} \frac{1}{\binom{n}{n_j}} q(n:n_j) p(\dots, \widehat{n_j}, \dots) \tag{9.1}$$

for any composition $(n_1, n_2, \dots, n_\ell)$ of $n$, which is a recursion analogous to (2.7). This allows an explicit formula,

$$p(n_1, n_2, \dots, n_\ell) = \sum_\sigma \frac{q(N_{\sigma(1)} : n_{\sigma(1)}) \cdots q(N_{\sigma(\ell)} : n_{\sigma(\ell)})}{\binom{n}{n_1, \dots, n_\ell}}, \tag{9.2}$$

where the sum is over all permutations $\sigma : [\ell] \to [\ell]$, and $N_{\sigma(j)} = n_{\sigma(j)} + \cdots + n_{\sigma(\ell)}$.

9.2. *Subordinator.* Exchangeability implies the existence of a nonnegative finite measure on [0, 1] such that

$$\Phi(b:k) = \binom{n}{k} \int_0^1 x^{k-1}(1-x)^{b-k} \Lambda(dx), \tag{9.3}$$

a representation to be compared with (2.10). The cumulative rate for some transition when $\Pi_n(t)$ has $b$ active blocks equals

$$\Phi(b) := \Phi(b:1) + \cdots + \Phi(b:b) = \int_0^1 \frac{1-(1-x)^b}{x} \Lambda(dx).$$

The last formula is an integral representation of a Bernstein function, hence the measure $\Lambda(dx)/x$ can be associated with some subordinator [16]. Explicitly, by de Finetti's theorem, there exists the limit proportion $S_t$ of integers in $[n]$ that comprise the active blocks of $\Pi^*_n(t)$ as $n \to \infty$. The process $(-\log(1-S_t), t \geq 0)$ is a subordinator with $S_0 = 0$ and distribution determined by

$$\mathbb{E}[(1-S_t)^\lambda] = e^{-t\Phi(\lambda)}, \qquad t \geq 0, \lambda \geq 0,$$

which is a version of the Lévy–Khintchine formula in the form of the Mellin transform. The subordinator has a drift if $\Lambda$ has an atom at 0.

Putting the blocks of $E_\infty$ in increasing order of their freezing times yields an ordered exchangeable partition with ECPF

$$p(n_1,\dots,n_\ell)=\prod_{j=1}^{\ell}\frac{q(N_j:n_j)}{\binom{N_j}{n_j}},$$

where $N_j:=n_j+\cdots+n_\ell$. The closed range of the process $(S_t)$ is the complement $U^c$ to the paintbox $U$ of the ordered partition $E_\infty$.

9.3. *Related Markov chains.*

9.3.1. *Transient.* For regenerative partitions, the analogue of $\mathrm{FM}_n$ introduced in Section 4 is the following. Let $q_\infty=\{q(b:k), 1\le k\le b<\infty\}$ be a decrement matrix. If there are $b$ active blocks in a partially frozen partition of $[n]$, then with probability $q(b:k)$, any $k$ of $b$ active blocks are chosen uniformly at random and merged into a single frozen block. Consistency translates as the recursion

$$\begin{aligned}q(b:k)={}&\frac{k+1}{b+1}q(b+1:k+1)+\frac{b+1-k}{b+1}q(b+1:k)\\&+\frac{1}{b+1}q(b+1:1)q(b:k)\end{aligned}\tag{9.4}$$

with $q(1:1)=1$, which leads to

$$q(b:k)=\Phi(b:k)/\Phi(b),\qquad (1\le k\le b<\infty),$$

where $\Phi$ has the integral representation (9.3) with some measure $\Lambda$, unique up to a positive multiple.

9.3.2. *Recurrent.* The analogue of the operation $\mathrm{SA}_n$ introduced in Section 5, acting on ordinary partitions of $[n]$, is the following [13]. Given a decrement matrix $q$, let $K_n$ follow $q(n:\cdot)$. Choose a value $k$ for $K_n$, then starting from some partition $\pi_n$ of $[n]$, sample $k$ balls from $\pi_n$ uniformly without replacement and then append a new box with these $k$ balls to the remaining partition of $n-k$ balls. According to an ordered version of the algorithm acting on ordered partitions, the balls are sampled from a totally ordered series of boxes and the newly created box is always arranged as the first box in the series.

In contrast to the $\mathrm{SA}_n$ operation, these Markov chains on partitions of $[n]$ are consistent under restrictions as $n$ varies. To see that *the operations* $\mathrm{SA}_n$ *are not consistent as n varies* [excluding the hook case $q(n:1)+q(n:n)\equiv 1$], fix $n>2$ and let $\pi_{n+1}$ be a partition having a singleton block $\{n+1\}$. There is a chance that some $2\le r\le n$ balls are sampled from $\pi_{n+1}$ and added in the box $\{n+1\}$. In this case, the restriction of $\mathrm{SA}_{n+1}$ to $[n]$ *creates* a novel nonsingleton box, which is not a legitimate option for $\mathrm{SA}_n$.

In [13], it was shown that the unique stationary $[n]$-partition is the one given by (9.2).

EXAMPLE. When

$$q(n:1) = \frac{n\rho}{1+n\rho}, \qquad q(n:n) = \frac{1}{1+n\rho},$$

the operation will create a new singleton block with probability $q(n:1)$ and will merge everything in one block with probability $q(n:n)$. So the stationary distributions will concentrate on hook partitions. The decrement matrix for this chain is the same as for $\mathrm{SA}_n$.

EXAMPLE. When

$$q(n:m) = \binom{n}{m} \frac{[\theta]_{n-m} m!}{[\theta+1]_{n-1} n},$$

with $\theta = 2\rho$, the invariant partition is Ewens', with parameter $\theta$. The decrement matrix for this chain is different from the one for $\mathrm{SA}_n$, which also leads to Ewens' distribution.

9.4. *Comparing recursions of EPPF.* Both (2.7) and (9.1) are forward recursions, which determine Möhle's partitions and regenerative partitions, respectively. They carry a striking parallel, except that the latter allows an explicit formula (9.2). In both cases, it is clear that each of these functions $p$ can be written as a linear combination of products of entries of the decrement matrix. To illustrate the similarity between the two recursions (2.7) and (9.1), we list the first few values of $p$ in terms of the decrement matrix $q$, first for Möhle's recursion (2.7):

$$\begin{aligned}
p(1) &= 1,\\
p(2) &= q(2:2),\\
p(1,1) &= q(2:1),\\
p(3) &= q(3:3) + q(3:2)q(2:2),\\
p(2,1) &= p(1,2) = \tfrac{1}{3}q(3:2)q(2:1) + \tfrac{1}{3}q(3:1)q(2:2),\\
p(1,1,1) &= q(3:1)q(2:1),\\
p(4) &= q(4:4) + q(4:3)q(2:2) + q(4:2)q(3:3)\\
&\quad + q(4:2)q(3:2)q(2:2),\\
p(3,1) &= p(1,3) = \tfrac{1}{4}q(4:3)q(2:1) + \tfrac{1}{6}q(4:2)q(3:2)q(2:1)\\
&\quad + \tfrac{1}{2}q(4:2)q(3:1)q(2:2) + \tfrac{1}{4}q(4:1)q(3:3)\\
&\quad + \tfrac{1}{12}q(4:1)q(3:2)q(2:2),\\
p(2,1,1) &= p(1,2,1) = p(1,1,2)\\
&= \tfrac{1}{6}q(4:2)q(3:1)q(2:1) + \tfrac{1}{6}q(4:1)q(3:2)q(2:1)
\end{aligned}$$

$$+\tfrac{1}{6}q(4:1)q(3:1)q(2:2),$$

$$p(1,1,1,1)=q(4:1)q(3:1)q(2:1).$$

The condition on $q_\infty$ equivalent to consistency of $p$ while $n$ varies has been described in Lemma 4.4.

Similarly, the first few values of the $p$ determined by a decrement matrix $q$ via the recursion (9.1) associated with a regenerative partition structure are

$$\begin{aligned}
p(1) &= 1,\\
p(2) &= q(2:2),\\
p(1,1) &= q(2:1),\\
p(3) &= q(3:3),\\
p(2,1) &= p(1,2)\\
&= \tfrac{1}{3}q(3:2)+\tfrac{1}{3}q(3:1)q(2:2),\\
p(1,1,1) &= q(3:1)q(2:1),\\
p(4) &= q(4:4),\\
p(3,1) &= p(1,3)\\
&= \tfrac{1}{4}q(4:3)+\tfrac{1}{4}q(4:1)q(3:3),\\
p(2,1,1) &= p(1,2,1)\\
&= p(1,1,2)=\tfrac{1}{6}q(4:2)q(2:1)\\
&\qquad +\tfrac{1}{6}q(4:1)q(3:2)+\tfrac{1}{6}q(4:1)q(3:1)q(2:2),\\
p(1,1,1,1) &= q(4:1)q(3:1)q(2:1).
\end{aligned}$$

The parallel condition on $q_\infty$ equivalent to consistency of $p$ while $n$ varies is described in (9.4).

Looking at these displays, both similarities and differences may be observed. In particular, the formulas for singleton partitions $(1,1,\dots,1)$ are identical. As is to be expected, the simpler recursion (9.1) for regenerative partitions generates simpler algebraic expressions than Möhle's recursion (2.7). See [13], equation (16), for the general formula for the shape function associated with (9.1).

9.5. *Comparing decrement matrices.* In [15], we found very similar recursions for entries of a decrement matrix which characterizes a regenerative composition structure, hence a regenerative partition structure in [13]. According to [15], Proposition 3.3, a nonnegative matrix $q$ is the decrement matrix of some regenerative composition structure if and only if $q(1:1)=1$ and (9.4) holds for $1\le k\le b$.

Comparing with Lemma 4.4 above, the difference from our recursions here is that we have a separate recursion for $q(b:1)$ and that we have an extra term

$$\frac{2}{b+1}q(b+1:2)q(b:k)$$

in the right-hand side of recursions for $q(b:k)$, $k \geq 2$. Both are backward recursions. For the purpose of illustration, supposing that we are given $q(4:k)$, $k = 1, 2, 3, 4$, the entries $q(b:\cdot)$ with $b \leq 3$ of decrement matrix for regenerative composition structure would be

$$\begin{aligned}
q(3:3) &= \frac{4q(4:4) + q(4:3)}{4 - q(4:1)}, \\
q(3:2) &= \frac{3q(4:3) + 2q(4:2)}{4 - q(4:1)}, \\
q(3:1) &= \frac{2q(4:2) + 3q(4:1)}{4 - q(4:1)}, \\
q(2:2) &= \frac{3q(3:3) + q(3:2)}{3 - q(3:1)} = \frac{6q(4:4) + 3q(4:3) + q(4:2)}{6 - 3q(4:1) - q(4:2)}, \\
q(2:1) &= \frac{2q(3:2) + 2q(3:1)}{3 - q(3:1)} = \frac{3q(4:3) + 4q(4:2) + 3q(4:1)}{6 - 3q(4:1) - q(4:2)}.
\end{aligned}$$

For the decrement matrix of the partition structure studied here, we have

$$\begin{aligned}
q(3:3) &= \frac{4q(4:4) + q(4:3)}{4 - q(4:1) - 2q(4:2)}, \\
q(3:2) &= \frac{3q(4:3) + 2q(4:2)}{4 - q(4:1) - 2q(4:2)}, \\
q(3:1) &= \frac{3q(4:1)}{4 - q(4:1) - 2q(4:2)}, \\
q(2:2) &= \frac{3q(3:3) + q(3:2)}{3 - q(3:1) - 2q(3:2)} = \frac{6q(4:4) + 3q(4:3) + q(4:2)}{6 - 3q(4:1) - 5q(4:2) - 3q(4:3)}, \\
q(2:1) &= \frac{2q(3:1)}{3 - q(3:1) - 2q(3:2)} = \frac{3q(4:1)}{6 - 3q(4:1) - 5q(4:2) - 3q(4:3)}.
\end{aligned}$$

9.6. *Comparison with Markovian fragmentations.* The theory of homogeneous and self-similar Markovian fragmentation processes due to Bertoin [2] is formulated much like the present theory of coalescents, in terms of consistent partition-valued processes. Ford ([11], Proposition 41) provides a sampling consistency condition for decrement matrices associated with discrete fragmentation processes which is an extremely close relative of our Lemma 4.4. The article [18] provides an integral representation for such decrement matrices, analogous to our results for the decrement matrices associated with regenerative partition structures and with Markovian coalescents, and embeds Ford's result in the broader context of continuous-time fragmentation processes and continuum random trees. A missing element of the fragmentation discussion is some way of deriving a partition structure by a recursion like (2.7) or (9.1). But we expect such a partition structure and an associated recursion may be associated with a suitably defined Markovian fragmentation with freeze, such as that introduced in [17].

## REFERENCES

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII—1983. Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. MR0883646

[2] BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes.* Cambridge Univ. Press. MR2253162

[3] BERTOIN, J. and GOLDSCHMIDT, C. (2004). Dual random fragmentation and coagulation and an application to the genealogy of Yule processes. In *Math. and Comp. Sci. III. Trends Math.* 295–308. Birkhäuser, Basel. MR2090520

[4] BOLTHAUSEN, E. and SZNITMAN, A. S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197** 247–276. MR1652734

[5] CANNINGS, C. (1974). The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. in Appl. Probab.* **6** 260–290. MR0343949

[6] DONG, R., GOLDSCHMIDT, C. and MARTIN, J. B. (2005). Coagulation-fragmentation duality, Poisson–Dirichlet distributions and random recursive trees. Available at http://front.math.ucdavis.edu/math.PR/0507591.

[7] DONNELLY, P. and JOYCE, P. (1991). Consistent ordered sampling distributions: characterization and convergence. *Adv. in Appl. Probab.* **23** 229–258. MR1104078

[8] DONNELLY, P. and TAVARÉ, S. (1986). The ages of alleles and a coalescent. *Adv. in Appl. Probab.* **18** 1–19. MR0827330

[9] EVANS, S. N. and PITMAN, J. (1998). Construction of Markovian coalescents. *Ann. Inst. H. Poincaré Probab. Statist.* **34** 339–383. MR1625867

[10] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biology* **3** 87–112. (Erratum, ibid. **3** 240; erratum, ibid. **3** 376.) MR0325177

[11] FORD, D. J. (2005). Probabilities on cladograms: introduction to the alpha model. Available at http://front.math.ucdavis.edu/math.PR/0511246.

[12] GNEDIN, A. (1997). The representation of composition structures. *Ann. Probab.* **25** 1437–1450. MR1457625

[13] GNEDIN, A. and PITMAN, J. (2004). Regenerative partition structures. *Electron. J. Combin.* **11** Research Paper 12 21. MR2120107

[14] GNEDIN, A. and PITMAN, J. (2005). Markov and self-similar composition structures. *Zapiski Nauchnych Seminarov POMI* **326** 59–84. Available at http://www.pdmi.ras.ru/znsl/2005/v326.html. MR2183216

[15] GNEDIN, A. and PITMAN, J. (2005). Regenerative composition structures. *Ann. Probab.* **33** 445–479. MR2122798

[16] GNEDIN, A. and PITMAN, J. (2006). Moments of convex distribution functions and completely alternating sequences. Available at http://front.math.ucdavis.edu/math.PR/0602091.

[17] GNEDIN, A. and YAKUBOVICH, Y. (2006). Recursive partition structures. *Ann. Probab.* **34** 2203–2218.

[18] HAAS, B., MIERMONT, G., PITMAN, J. and WINKEL, M. (2006). Asymptotics of discrete fragmentation trees and applications to phylogenetic models. Available at http://front.math.ucdavis.edu/math.PR/0604350.

[19] KINGMAN, J. F. C. (1978). The representation of partition structures. *J. London Math. Soc. (2)* **18** 374–380. MR0509954

[20] KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034

[21] KINGMAN, J. F. C. (1982). Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics* (*Rome*, *1981*) 97–112. North-Holland, Amsterdam. MR0675968

[22] KINGMAN, J. F. C. (1982). On the genealogy of large populations. *J. Appl. Probab.* **19A** 27–43. MR0633178

[23] MÖHLE, M. (2006). On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12** 35–53. MR2202319

[24] MÖHLE, M. (2006). On a class of non-regenerative sampling distributions. *Combin. Probab. Comput.* **16** 435–444.

[25] MÖHLE, M. and SAGITOV, S. (2001). A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29** 1547–1562. MR1880231

[26] MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.* **54** 60–71. MR0127989

[27] NORDBORG, M. (2001). Coalescent theory. In *Handbook of Statistical Genetics* (D. J. Balding et al., eds.) 179–208. Wiley, New York.

[28] PITMAN, J. (2006). *Combinatorial stochastic processes*. *École d'Été de Probabilités de Saint-Flour XXXII—2002*. *Lecture Notes Math.* **1875**. MR2245368

[29] PITMAN, J. (1999). Coalescents with multiple collisions. *Ann. Probab.* **27** 1870–1902. MR1742892

[30] SAGITOV, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36** 1116–1125. MR1742154

[31] SAGITOV, S. (2003). Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40** 839–854. MR2012671

[32] SCHWEINSBERG, J. (2000). A necessary and sufficient condition for the $\Lambda$-coalescent to come down from infinity. *Electron. Comm. Probab.* **5** 1–11. MR1736720

[33] SCHWEINSBERG, J. (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5** paper 12 50. MR1781024

[34] TAVARÉ, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoret. Population Biol.* **26** 119–164. MR0770050

[35] WATTERSON, G. A. (1976). Reversibility and the age of an allele. *Theoret. Population Biol.* **10** 239–253. MR0475994

[36] WATTERSON, G. A. (1984). Lines of descent and the coalescent. *Theoret. Population Biol.* **26** 77–92. MR0760232

[37] YOUNG, J. E. (1995). Partition-valued stochastic processes with applications. Ph.D. thesis, Univ. California, Berkeley.

R. DONG
J. PITMAN
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
367 EVANS HALL # 3860
BERKELEY, CALIFORNIA 94720-3860
USA
E-MAIL: ruidong@stat.berkeley.edu
pitman@stat.berkeley.edu

A. GNEDIN
DEPARTMENT OF MATHEMATICS
UTRECHT UNIVERSITY
PO BOX 80.010
3508 TA UTRECHT
THE NETHERLANDS
E-MAIL: gnedin@math.uu.nl