

# THE COALESCENT POINT PROCESS OF BRANCHING TREES

BY AMAURY LAMBERT<sup>1</sup> AND LEA POPOVIC<sup>2</sup>

*UPMC Univ Paris 06 and Concordia University*

We define a doubly infinite, monotone labeling of Bienaymé–Galton–Watson (BGW) genealogies. The genealogy of the current generation backwards in time is uniquely determined by the coalescent point process  $(A_i; i \geq 1)$ , where  $A_i$  is the coalescence time between individuals  $i$  and  $i + 1$ . There is a Markov process of point measures  $(B_i; i \geq 1)$  keeping track of more ancestral relationships, such that  $A_i$  is also the first point mass of  $B_i$ .

This process of point measures is also closely related to an inhomogeneous spine decomposition of the lineage of the first surviving particle in generation  $h$  in a planar BGW tree conditioned to survive  $h$  generations. The decomposition involves a point measure  $\rho$  storing the number of subtrees on the right-hand side of the spine. Under appropriate conditions, we prove convergence of this point measure to a point measure on  $\mathbb{R}_+$  associated with the limiting continuous-state branching (CSB) process. We prove the associated invariance principle for the coalescent point process, after we discretize the limiting CSB population by considering only points with coalescence times greater than  $\varepsilon$ .

The limiting coalescent point process  $(B_i^\varepsilon; i \geq 1)$  is the sequence of depths greater than  $\varepsilon$  of the excursions of the height process below some fixed level. In the diffusion case, there are no multiple ancestries and (it is known that) the coalescent point process is a Poisson point process with an explicit intensity measure. We prove that in the general case the coalescent process with multiplicities  $(B_i^\varepsilon; i \geq 1)$  is a Markov chain of point masses and we give an explicit formula for its transition function.

The paper ends with two applications in the discrete case. Our results show that the sequence of  $A_i$ 's are i.i.d. when the offspring distribution is linear fractional. Also, the law of Yaglom's quasi-stationary population size for subcritical BGW processes is disintegrated with respect to the time to most recent common ancestor of the whole population.

**1. Introduction.** The idea of describing the backward genealogy of a population is ubiquitous in population genetics. The most popular piece of work on the

---

Received January 2011; revised October 2011.

<sup>1</sup>Supported by the Projects MAEV 06-BLAN-3 146282 and MANEGE 09-BLAN-0215 of ANR (French national research agency).

<sup>2</sup>Supported by the University Faculty award and Discovery grant of NSERC (Natural sciences and engineering council of Canada).

*MSC2010 subject classifications.* Primary 60J80; secondary 60G55, 60G57, 60J10, 60J85, 60J27, 92D10, 92D25.

*Key words and phrases.* Coalescent point process, branching process, excursion, continuous-state branching process, Poisson point process, height process, Feller diffusion, linear-fractional distribution, quasi-stationary distribution, multiple ancestry.

subject is certainly [17], where the *standard coalescent* is introduced, and shown to describe the genealogy of a finite sample from a population with large but constant population size. Coalescent processes for branching processes cannot be characterized in the same way, since, for example, they are not generally Markov as time goes backwards, although for stable continuous-state branching (CSB) processes the genealogy can be seen as the time-change of a Markovian coalescent [6].

The present paper relies on the initial works [1, 26], which focused on the coalescent point process for critical birth–death processes and the limiting Feller diffusion. They have been extended to noncritical birth–death processes [13] and more generally to homogeneous, binary Crump–Mode–Jagers processes [23]. In all these references, simultaneous births were not allowed, since then the genealogical process would have to keep memory of the multiplicity of all common offspring of an ancestor. The problem of sampling a branching population or a coalescent point process has received some attention [19, 22, 28], but no consistent sampling in the standard coalescent has been given so far (except Bernoulli sampling of leaves). Our goal was to define a coalescent point process for arbitrary branching processes, that is both simple to describe in terms of its law, and allows for finite sampling in a consistent way. That is, the coalescent process for samples of size  $n \geq 1$  are all embedded in the same object.

A different way of characterizing the genealogy of a branching population alive at some fixed time is with a *reduced tree*, first studied in [10] and generalized in [8], Section 2.7, to CSB processes. To construct it, a starting date 0 and a finishing date  $T$  are specified, and a reduced branching tree started at 0 and conditioned to be alive at  $T$  is defined by erasing the points without alive descendants at time  $T$ . Instead of directly displaying the coalescence times as a sequence running over the current population size (as is our goal), this approach characterizes the transition probabilities of the reduced branching process by tracking the population size in time with an inhomogeneous Markov process on  $[0, T]$  taking values in the set of integers for all times in  $[0, T]$ . Unfortunately, this construction does not allow for a consistent way of sampling the individuals alive at time  $T$ .

We use a different approach and construct a coalescent process with multiplicities for the genealogy of some random population, when the forward time genealogy is produced by a general branching process, either discrete or continuous-state. Our main goal is to give a simple representation for this process, and describe its law in a manner that would be easy to use in applications.

We are interested in having an arbitrarily large population at the present time, that we think of as generation 0, arising from a general branching process, originating at an unspecified arbitrarily large time in the past. In order to keep track of genealogies of individuals in the present population, for discrete Bienaymé–Galton–Watson (BGW) branching processes we use a representation that labels individuals at each generation in such a way that lines of descent of the population at any time do not intersect (see Figure 1). This leads to a monotone planar embedding of BGW trees such that each individual is represented by  $(n, i)$  where  $n \in \mathbb{Z}$

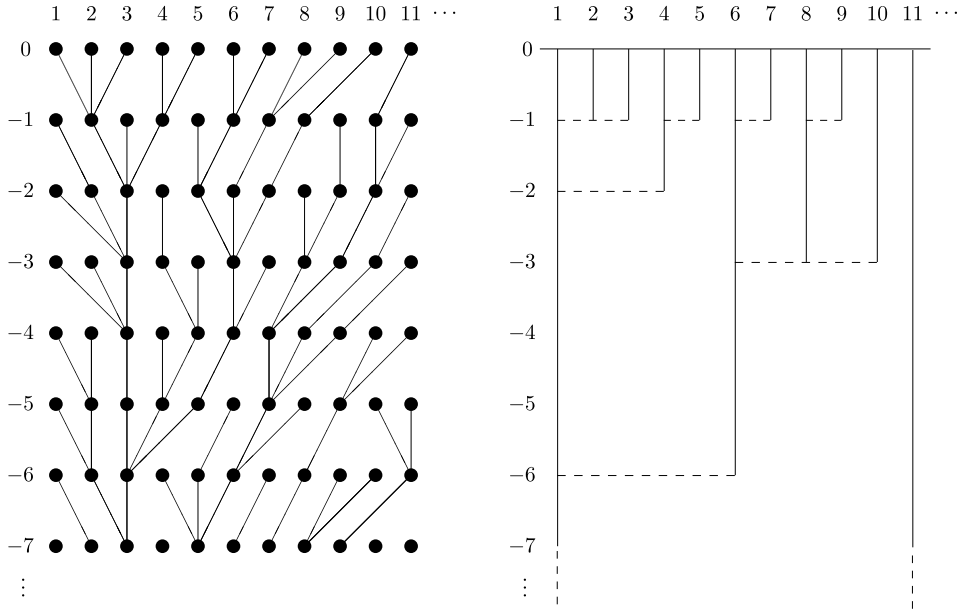


FIG. 1. A doubly infinite embedding of quasi-stationary Bienaymé–Galton–Watson genealogies. Horizontal labels are individuals and vertical labels are generations. Left panel: the complete embedding of an infinite BGW tree. Right panel: coalescence times between consecutive individuals of generation 0. Here, the coalescent point process takes the value  $(1, 1, 2, 1, 6, 1, 3, 1, 3, \dots)$ .

is the generation number and  $i \in \mathbb{N}$  is the individual’s index in the generation. We consider BGW trees which are doubly infinite, allowing the number of individuals alive at present time to be arbitrarily large, and considering an infinite number of generations for their ancestry back in time. This monotone representation can also be extended to the case of continuous-state branching processes (later called CSB processes), in terms of a linearly ordered planar embedding of the associated genealogical  $\mathbb{R}$ -trees. This is in some way implicit in many recent works: our embedding is the exact discrete analogue of the flow of subordinators defined in [4]; also, in the genealogy of CSB processes defined in [24], the labeling of individuals at a fixed generation is analogous to the successive visits of the height process at a given level.

From our discrete embedding of the population, a natural consequence is that *coalescence times*, or times backwards to the most recent common ancestor, between individuals  $k$  and  $\ell$  in generation 0, are represented by the maximum of the values  $A_i, i = k, \dots, \ell - 1$ , where  $A_i$  is the coalescence time between individuals  $i$  and  $i + 1$ . The genealogy back in time of the present population is then uniquely determined by the process  $(A_i; i \geq 1)$ , which we call the *branch lengths* of the coalescent point process. In the continuous-state branching process setting, the  $A_i$ ’s are given by depths of the excursions of the height process below the level representing the height of individuals in the present population.

In general, the process  $(A_i; i \geq 1)$  is not Markovian and its law is difficult to characterize. Our key strategy is to keep track of multiple ancestries to get a Markov process, thus constructing a *coalescent point process with multiplicities*  $(B_i; i \geq 1)$ . Intuitively, for each  $i \geq 1$ ,  $B_i$  encodes the relationship of the individual  $i + 1$  to the infinite spine of the first present individual, by recording the nested sequence of subtrees that form that ancestral lineage linking  $i + 1$  to the spine. The value of  $B_i$  is a point mass measure, where each point mass encodes one of these nested subtrees: the level of the point mass records the level back into the past at which this subtree originated, while the multiplicity of the point mass records the number of subtrees with descendants in individuals  $\{i + 1, i + 2, \dots\}$  emanating at that level which are embedded on the right-hand side of the ancestral link of  $i + 1$  to the spine.

Formally, for both BGW and the CSB population, we proceed as follows. We define a process with values in the integer-valued measures on  $\mathbb{N}$  or  $\mathbb{R}_+$  (BGW or CSB, resp.) in a recursive manner, in such a way that  $(B_j; 1 \leq j \leq i)$  will give a complete record of the ancestral relationship of individuals  $1 \leq j \leq i$ . Start with the left-most individual in the present population,  $i = 1$ , and let  $B_1$  have a point mass at  $n$ , where  $-n$  is the generation of the last common ancestor of individuals 1 and 2, and have multiplicity  $B_1(\{n\})$  equal to the number of times this ancestor will also appear as the last common ancestor for individuals ahead, with index  $i \geq 1$ . We then proceed recursively, so that at individual  $i$  we take the point masses in  $B_{i-1}$ ; we first make an adjustment in order to reflect the change in multiplicities due to the fact that the lineage of individual  $i$  is no longer considered to be an individual ahead of the current one. We then add a record for the last common ancestor of individuals  $i$  and  $i + 1$ . If this is an ancestor that has already been recorded in  $B_{i-1}$ , we just let  $B_i$  be the updated value of  $B_{i-1}$ . Otherwise, we let  $B_i$  be the updated  $B_{i-1}$  plus a new point mass at  $n$ , if  $-n$  is the generation of this new last common ancestor, whose multiplicity  $B_i(\{n\})$  is the number of times this new ancestor will appear as the last common ancestor for individuals ahead, with index  $\geq i + 1$  (e.g., in Figure 1 we will have  $B_1 = 2\delta_1$ ,  $B_2 = \delta_1$ ,  $B_3 = \delta_2$ ,  $B_4 = \delta_1$ ,  $B_5 = \delta_6$ ,  $B_6 = \delta_1$ ,  $B_7 = 2\delta_3$ ,  $B_8 = \delta_1 + \delta_3$ ,  $B_9 = \delta_3$ ).

Once we have constructed the coalescent point process with multiplicities (in either the BGW or CSB case), we will show that  $A_i$  can be recovered as the location of the nonzero point mass in  $B_i$  with the smallest level, that is,  $A_i = \inf\{n : B_i(\{n\}) \neq 0\}$  (e.g., in Figure 1 we have  $A_1 = 1$ ,  $A_2 = 1$ ,  $A_3 = 2$ ,  $A_4 = 1$ ,  $A_5 = 6$ ,  $A_6 = 1$ ,  $A_7 = 3$ ,  $A_8 = 1$ ,  $A_9 = 3$ ). More importantly, we prove that  $(B_i; i \geq 1)$  is a Markov process, and that, when going from individual  $i$  to  $i + 1$ , the transitions decrease the multiplicity of the point mass of  $B_i$  at level  $\inf\{n : B_i(\{n\}) \neq 0\}$  by 1, and that, with a specified probability, a new random point mass is added at a random level that must be smaller than the smallest nonzero level in the updated version of  $B_i$ .

In order to use this construction for both BGW and the CSB population, our take on what constitutes the sequence of present individuals has to be different for the

continuous CSB population from the simple one for the discrete BGW population. Since in the CSB case the present population size is not discrete, and there is an accumulation of immediate ancestors at times arbitrarily close to the present time, we have to discretize the present population by considering only the individuals whose last common ancestor occurs at a time at least an  $\varepsilon$  amount below the present time, for an arbitrary  $\varepsilon > 0$ . We will later show that we can obtain the law of the coalescent point process with multiplicities for the CSB population as a limit of a sequence of appropriately rescaled coalescent point processes with multiplicities for the BGW population for which we also use the same discretization process of the present population. At first it may seem surprising that these coalescents with multiplicities are Markov processes over the set of all or the  $\varepsilon$ -discretized set (BGW and CSB case, resp.) individuals at present time. Below we intuitively explain why this is the case by describing the two approaches for constructing them.

In the discrete case, we start by giving a related, easier to define process  $(D_i; i \geq 1)$ , taking values in the integer-valued sequences, whose first nonzero term is also at level  $A_i$ . For each  $i$ , the sequence  $(D_i(n), n \geq 1)$  gives the number of younger offshoots at generation  $-n$  embedded on the right-hand side of the ancestor of  $i$ . The trees sprouting from the younger offshoots are independent, and the law of a tree sprouting from a younger offshoot at generation  $-n$  has the law of the BGW tree conditioned to survive at least  $n - 1$  generations. It turns out that  $(D_i; i \geq 1)$  is Markov, and we construct  $(B_i; i \geq 1)$  from it, show that it is Markov as well and give its transition law. In order to be able to deal with the conditioning of younger offshoots in a way that allows us later to pass to the limit, we introduce an integer-valued measure  $\rho$  that takes the ancestor of individual 1 at generation  $-n$  and records as  $\rho(\{n\}) = \rho_n$  the number of all of its younger offshoots embedded on the right-hand side, and we call it the *great-aunt measure*. This measure gives a spine decomposition of the first survivor (individual  $i = 1$ ) in such a way that the law of the trees sprouting from the younger offshoots are still independent, but are no longer conditioned on survival.

In the continuous-state case, the great-aunt measure will be a measure  $\rho^0$  on  $\mathbb{R}_+$ , and we will be able to define the genealogy thanks to independent CSB processes starting from the masses of the atoms of  $\rho^0$ . In the subcritical and critical cases, this can be done using a single path of a Lévy process with no negative jumps and Laplace exponent  $\psi$ . In the supercritical case a concatenation of excursion paths will have to be used. Using the continuous great-aunt measure  $\rho^0$ , we characterize the genealogy of an infinite CSB tree with branching mechanism  $\psi$  via the height function  $H^*$  whose value at an individual in the population can be decomposed into the level on the infinite spine at which the subtree containing this individual branches off, and the relative height of this individual within its subtree. We construct  $(B_i^\varepsilon; i \geq 1)$  from the height process  $H^*$ . Discretizing the population by considering only the points whose coalescence times are greater than some fixed  $\varepsilon > 0$ , translates into considering only the excursions of  $H^*$  from level 0

with a depth greater than  $-\varepsilon$ . From these excursions we can obtain the process  $((A_i^\varepsilon, N_i^\varepsilon); i \geq 1)$  where  $A_i^\varepsilon$  is the depth of the  $i$ th such excursion and  $N_i^\varepsilon$  is the number of future excursions with the exact same depth. It turns out that a specific functional of this process is Markov, and we construct  $(B_i^\varepsilon; i \geq 1)$  from it, show that it is Markov as well and give its transition law.

Now, in order to prove convergence in law of the discretized version of  $(B_i; i \geq 1)$  to  $(B_i^\varepsilon; i \geq 1)$ , we make use of the great-aunt measures  $\rho$  and  $\rho^0$  describing the spine decomposition of the first surviving individual. In the discrete case, the spine decomposition  $(\rho(\{n\}); 0 \leq n \leq h)$  truncated at level  $h$  has the same law as the spine decomposition of a planar embedding of a BGW tree conditioned on surviving up to generation  $h$ , where the first survivor is defined in the usual depth-first search order (see Figure 2). Using a well-known random walk representation of BGW trees [3], we then prove (under usual conditions ensuring the convergence of the random walk to a spectrally positive Lévy process  $X$  with Laplace exponent  $\psi$ ) the convergence (in the vague topology) of the great-aunt measure  $\rho$

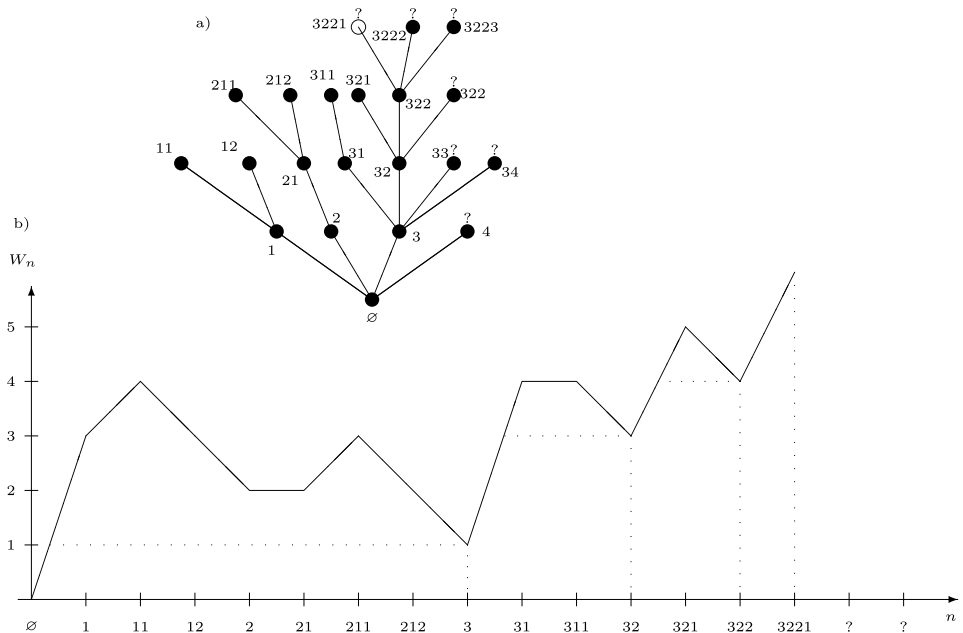


FIG. 2. (a) A Bienaymé–Galton–Watson tree conditioned to have alive individuals at generation 4. The empty circle is the first such individual (3221) in the lexicographical order. Descendants of individuals with greater rank are only indicated by a question mark; (b) the associated random walk  $W$  killed after the visit  $\sigma_4$  of the first individual  $x_4 = 3221$  with height 4. Records of the future infimum at  $\sigma_4$  are shown by dotted lines. The record times are exactly the visits to ancestors 322, 32, 3 and  $\emptyset$  of 3221. The pairs of overshoots and undershoots of  $W$  across the future infimum at those times (in this order) are  $(\alpha_1, \rho_1) = (0, 2)$ ,  $(\alpha_2, \rho_2) = (1, 1)$ ,  $(\alpha_3, \rho_3) = (1, 2)$  and  $(\alpha_4, \rho_4) = (2, 1)$ .

to the measure  $\rho^0$  on  $\mathbb{R}_+$  defined by

$$\rho^0(dx) := \beta dx + \sum_{t: \Delta_t > 0} \Delta_t \delta_t(dx),$$

where  $\beta$  is the Gaussian coefficient of  $X$ , and  $(t, \Delta_t)$  is a Poisson point measure with intensity

$$\left( e^{v(t)r} \int_{(r, \infty)} e^{-v(t)z} \Lambda(dz) \right) dt dr, \quad r > 0,$$

where  $\Lambda$  is the Lévy measure of  $X$ , and  $v$  the inverse of the decreasing function  $\lambda \mapsto \int_{[\lambda, +\infty)} du/\psi(u)$ . This is exactly the same measure  $\rho^0$  we obtain from the spine decomposition in the continuous case. Since the transition law of  $(B_i; i \geq 1)$  can be represented as a functional of  $\rho$  and the transition law of  $(B_i^\varepsilon; i \geq 1)$  as a functional of  $\rho^0$ , this will lead to our claim.

Finally, in the very last section, we give two simple applications of our results in the discrete case. First, we prove that in the linear-fractional case, the coalescent point process is a sequence of i.i.d. random variables. Related results can be found in [27]. Second, in the subcritical case, we use the monotone embedding to display the law, in quasi-stationary state, of the total population size (Yaglom’s limit) jointly with the time to most recent common ancestor of this whole population.

**2. Doubly infinite embedding and the coalescent point process.**

2.1. *The discrete model.* We will start off with a monotone planar embedding of an infinitely old BGW process with arbitrarily large population size. Let  $(\xi(n, i))$  be a doubly indexed sequence of integers, where  $n$  is indexed by  $\mathbb{Z} = \{\dots, 0, -1, -2, -3, \dots\}$  and  $i$  is indexed by  $\mathbb{N} = \{1, 2, 3, \dots\}$ . The index pair  $(n, i)$  represents the  $i$ th individual in generation  $n$ , and  $\xi(n, i)$  provides the number of offspring of this individual.

We endow the populations with the following genealogy. Individual  $(n, i)$  has mother  $(n - 1, j)$  if

$$\sum_{k=1}^{j-1} \xi(n - 1, k) < i \leq \sum_{k=1}^j \xi(n - 1, k).$$

From now on, we focus on the ancestry of the population at time 0, that we will call *standing population*, and we let  $\alpha_i(n)$  denote the index of the ancestor of individual  $(0, i)$  in generation  $-n$ . In particular,  $\alpha_i(1) := \min\{j \geq 1 : \sum_{k=1}^j \xi(-1, k) \geq i\}$ . Our main goal is to describe the law of the times of coalescence  $C_{i,j}$  of individuals  $(0, i)$  and  $(0, j)$ , that is,

$$C_{i,j} := \min\{n \geq 1 : \alpha_i(n) = \alpha_j(n)\},$$

where it is understood that  $\min \emptyset = +\infty$ . Defining

$$A_i := C_{i,i+1},$$

it is easily seen that by construction, for any  $i \leq j$ ,  $C_{i,j} = \max\{A_i, A_{i+1}, \dots, A_{j-1}\}$ . Thus, the sequence  $A_1, A_2, \dots$  contains all the information about the (un-labeled) genealogy of the current population and is called *coalescent point process*, as in [26] (see Figure 1).

We assume that there is a random variable  $\xi$  with values in  $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$  and probability generating function (p.g.f.)  $f$ , such that all random variables (r.v.s)  $\xi(n, i)$  are i.i.d. and distributed as  $\xi$ . As a consequence, if  $Z^{(n,i)}(k)$  denotes the number of descendants of  $(n, i)$  at generation  $n + k$ , then the processes  $(Z^{(n,i)}(k); k \geq 0)$  are identically distributed Bienaymé–Galton–Watson (BGW) processes starting from 1. We will refer to  $Z$  as some generic BGW process with this distribution and we will denote by  $f_n$  the  $n$ th iterate of  $f$  and by  $p_n = 1 - f_n(0)$  the probability that  $Z_n \neq 0$ . We need some extra notation in this setting.

We define  $\zeta_n$  as the number of individuals at generation 1 having alive descendants at generation  $n$ . In particular,  $\zeta_n$  has the law of

$$\sum_{k=1}^{\xi} \epsilon_k,$$

where the  $\epsilon_k$  are i.i.d. Bernoulli r.v.s with success probability  $p_{n-1}$ , independent of  $\xi$ . We also define  $\zeta'_n$  as the r.v. distributed as  $\zeta_n - 1$  conditional on  $\zeta_n \neq 0$ .

*2.2. The coalescent point process: Main results.* Recall that  $A_i$  denotes the time of coalescence of individuals  $(0, i)$  and  $(0, i + 1)$ , or  $i$ th *branch length of the coalescent*. To describe the law of  $(A_i; i \geq 1)$  we need an additional process.

Let  $D_i(n)$  be the number of daughters of  $(-n, \alpha_i(n))$ , distinct from  $(-n + 1, \alpha_i(n - 1))$ , having descendants in  $\{(0, j); j \geq i\}$ . In other words,  $D_i(n)$  is the number of younger surviving offshoots of  $(-n, \alpha_i(n))$  not counting the lineage of  $(0, i)$  itself. Letting

$$\mathcal{D}_i(n) := \{\text{daughters of } (-n, \alpha_i(n)) \text{ with descendants in } \{(0, j); j \geq i\}\},$$

we have

$$D_i(n) = \#\mathcal{D}_i(n) - 1.$$

We set  $D_0(n) := 0$  for all  $n$ . We now provide the law of this process and its relationship to  $(A_i; i \geq 1)$ . We also set  $A_0 := +\infty$ , which is in agreement with  $\min \emptyset = +\infty$ .

**THEOREM 2.1.** *The  $i$ th branch length is a simple functional of  $(D_i; i \geq 0)$*

$$A_i = \min\{n \geq 1 : D_i(n) \neq 0\}.$$



In addition, the sequence-valued chain  $(D_i; i \geq 0)$  is a Markov chain started at the null sequence. For any sequence of nonnegative integers  $(d_n)_{n \geq 0}$ , the law of  $D_{i+1}$  conditionally given  $D_i(n) = d_n$  for all  $n$ , is given by the following transition. We have  $D_{i+1}(n) = d_n$  for all  $n > A_i$ ,  $D_{i+1}(A_i) = d_{A_i} - 1$  and the r.v.s  $D_{i+1}(n)$ ,  $1 \leq n < A_i$ , are independent, distributed as  $\zeta'_n$ . In particular, the law of  $A_1$  is given by

$$\begin{aligned} \mathbb{P}(A_1 > n) &= \prod_{k=1}^n \mathbb{P}(\zeta'_k = 0) = \frac{1}{1 - f_n(0)} \prod_{k=1}^n f'(f_{k-1}(0)) \\ &= \frac{f'_n(0)}{1 - f_n(0)} = \mathbb{P}(Z_n = 1 \mid Z_n \neq 0). \end{aligned}$$

PROOF. The following series of equivalences proves that  $A_i$  is the level of the first nonzero term of the sequence  $D_i$ .

$$\begin{aligned} A_i > n &\Leftrightarrow \forall k \leq n, \alpha_i(k) \neq \alpha_{i+1}(k) \\ &\Leftrightarrow \forall k \leq n, \forall j > i, \alpha_i(k) \neq \alpha_j(k) \\ &\Leftrightarrow \forall k \leq n, (-k, \alpha_i(k)) \text{ has no descendants in } \{(0, j) : j > i\} \\ &\Leftrightarrow \forall k \leq n, D_i(k) = 0. \end{aligned}$$

Thanks to this last result, we get

$$\alpha_i(A_i) = \alpha_{i+1}(A_i) \quad \text{and} \quad \alpha_i(A_i - 1) \neq \alpha_{i+1}(A_i - 1).$$

In particular,  $(-A_i + 1, \alpha_i(A_i - 1))$  has no descendants in  $\{(0, j) : j \geq i + 1\}$ , so that

$$\mathcal{D}_{i+1}(A_i) = \mathcal{D}_i(A_i) \setminus \{(-A_i + 1, \alpha_i(A_i - 1))\}$$

and  $\#\mathcal{D}_{i+1}(A_i) = \#\mathcal{D}_i(A_i) - 1$ , that is,  $D_{i+1}(A_i) = D_i(A_i) - 1$ .

Let us deal with the case  $n > A_i$ . By definition of  $A_i$ ,  $\alpha_i(n) = \alpha_{i+1}(n)$  for any  $n \geq A_i$ . As a consequence, for any  $n > A_i$ , each daughter of  $(-n, \alpha_{i+1}(n))$  has descendants in  $\{(0, j) : j \geq i + 1\}$  iff she has descendants in  $\{(0, j) : j \geq i\}$ . In other words,  $\mathcal{D}_{i+1}(n) = \mathcal{D}_i(n)$  and  $D_{i+1}(n) = D_i(n)$ .

Now we deal with the case  $n < A_i$ . Set

$$E_i := \{D_j(n) = d_{j,n}, \forall n \geq 1, j \leq i\},$$

where the  $(d_{j,n})_{n \geq 1, j \leq i}$  are fixed integer numbers, and let  $A_i$  be the value of the coalescence time of  $i$  and  $i + 1$  conditional on  $E_i$ , that is,  $A_i = \min\{n \geq 1 : d_{i,n} \neq 0\}$ . Now let  $\mathcal{T}(n, i)$  be the tree descending from  $(-n, \alpha_i(n))$  and set

$$I(n, i) := \min\{j \leq i : \alpha_j(n) = \alpha_i(n)\}.$$

Observe that the (unlabeled) tree  $\mathcal{T}(n, i)$  has the law of a BGW tree conditioned on having at least one descendant at generation 0. Now, because the event  $E_i$  only

concerns the descendants of daughters of ancestors of  $(-A_i + 1, a_{i+1}(A_i - 1))$ , the law of  $\mathcal{T}(A_i - 1, i + 1)$  conditional on  $E_i$  is still the law of a BGW tree conditioned on having at least one descendant at generation 0.

Also notice that conditional on  $E_i$ ,  $I(A_i - 1, i + 1) = i + 1$ . As a consequence,  $I(n, i + 1) = i + 1$  for any  $n < A_i$ , so that for any  $n < A_i$ , conditional on  $E_i$ ,

$$\mathcal{D}_{i+1}(n) := \{\text{daughters of } (-n, a_{i+1}(n)) \text{ with descendants at generation } 0\}.$$

The result follows recalling the law of  $\mathcal{T}(A_i - 1, i + 1)$  conditional on  $E_i$ .  $\square$

Recall that  $D_0$  is a null sequence,  $A_0 = +\infty$ , and that  $D_1(n) \stackrel{d}{=} \zeta'(n)$  for  $n \geq 1$ . This infinite sequence of values in  $\{D_1(n), n \geq 1\}$  contains information on the ancestral relationship of the individual  $(0, 1)$  and an arbitrarily large number of individuals in the standing population going back arbitrarily far into the past. Using the process  $(D_i; i \geq 1)$  we next define a sequence  $(B_i; i \geq 1)$  of finite point mass measures which contains the minimal amount of information needed to reconstruct  $A_1, A_2, \dots$  while remaining Markov. We do this for two reasons. First, if we are only interested in the ancestral relationship of finitely many individuals in the standing population, there is no need to keep track of an infinite sequence of values. Second, when we consider a rescaled limit of the BGW process to a CSB process, we will need to work with a sparser representation.

The main distinction between these two processes is that while  $D_i$  contains information on the ancestral relationship of  $(0, i)$  and  $(0, j)$  for both  $1 \leq j \leq i + 1$  and  $j > i + 1$ ,  $B_i$  will only contain information about the ancestral relationship of  $(0, i)$  and  $(0, j)$  for  $1 \leq j \leq i + 1$ . In other words, if, say,  $\max\{A_1, \dots, A_i\} = n$ , then  $B_i(m) = 0$  for all  $m > n$ . Moreover,  $B_i$  is defined directly from  $D_i$  by letting  $B_i(\{m\}) = D_i(m)$  for all  $m \leq n$ . In particular,  $B_1$  will have a single nonzero point mass at level  $A_1$ , and  $B_1(\{A_1\}) = D_1(A_1)$ .

We are now ready to define  $(B_i; i \geq 1)$  which we call the *coalescent point process with multiplicities*. Let  $(B_i; i \geq 0)$  be a sequence of finite point measures, started at the null measure, defined from  $(D_i; i \geq 0)$  recursively as follows. For any point measure  $b = \sum_{n \geq 1} b_n \delta_n$ , let  $\mathfrak{s}(b)$  denote the minimum of the support of  $b$ , that is,

$$\mathfrak{s}(b) := \min\{n \geq 1 : b_n \neq 0\}$$

with the convention that  $\mathfrak{s}(b) = +\infty$  if  $b$  is the null measure. If  $B_i = \sum_{n \geq 1} b_{n,i} \delta_n$  for some  $\{b_{n,i}\}_{n \geq 1} \in \mathbb{N}$ , let

$$a_{1,i} := \mathfrak{s}(B_i), \quad B_i^* := B_i - \delta_{a_{1,i}}, \quad a_{1,i}^* := \mathfrak{s}(B_i^*).$$

Then, define

$$B_{i+1} := \begin{cases} B_i^* + D_{i+1}(A_{i+1})\delta_{A_{i+1}}, & \text{if } A_{i+1} < a_{1,i}^* \text{ and } A_{i+1} \neq a_{1,i}, \\ B_i^*, & \text{otherwise.} \end{cases}$$

Note that by Theorem 2.1 we have  $D_{i+1}(A_i) = D_i(A_i) - 1$ , so by this definition  $B_{i+1} = \sum_{n \geq 1} b_{n,i+1} \delta_n$ , where  $\{b_{n,i+1}\}_{n \geq 1} \in \mathbb{N}$  satisfies

$$b_{n,i+1} := \begin{cases} D_{i+1}(n), & \text{if } n \text{ is such that } b_{n,i} \neq 0, \\ D_{i+1}(A_{i+1}), & \text{if } n = A_{i+1} \text{ and } A_{i+1} < a_{1,i}^* \text{ and } A_{i+1} \neq a_{1,i}, \\ 0, & \text{for all other } n. \end{cases}$$

Roughly,  $(B_i; i \geq 1)$  records the ancestral information in the following way.  $B_0$  is a null measure, and  $B_1$  will contain a single point mass  $B_1 = D_1(A_1) \delta_{A_1}$  recording the coalescence time  $A_1$  for individuals  $(0, 1)$  and  $(0, 2)$  in the location of its point mass. Since the last common ancestor of  $(0, 1)$  and  $(0, 2)$  may also be the last common ancestor of  $(0, 1)$ ,  $(0, 2)$  and some other individual  $(0, j)$  for  $j \geq 2$ , the multiplicity of this point mass will record the number of its future appearances in the coalescent point process. Recursively in every step, say  $i + 1$ , this record of point masses will need to be updated from  $B_i$  to  $B_i^*$ , since by moving one individual to the right we are no longer recording the last common ancestor of two previous individuals, and the number of future appearances of their last common ancestor by definition goes down by 1. In addition, at step  $i + 1$  we also need to record the coalescence time for the last common ancestor of  $(0, i + 1)$  and  $(0, i + 2)$  with the number of its future appearances. This is done by taking the updated version  $B_i^*$  and adding a new point mass  $D_{i+1}(A_{i+1}) \delta_{A_{i+1}}$  to create  $B_{i+1}$ . Because of the monotone embedding of the BGW as a planar tree, it is not possible for the level  $A_{i+1}$  of this new point mass to be greater than any of the common ancestors of  $(0, j)$  and  $(0, k)$  for  $1 \leq j < k \leq i + 1$  unless the multiplicities of these ancestors are depleted at this step and no longer appear in the updated  $B_i^*$  (so  $A_{i+1} \leq a_{1,i}^*$ ) as will be seen in the proof of the next theorem. Moreover, if the last common ancestor of  $(0, i + 1)$  and  $(0, i + 2)$  is the last common ancestor of  $(0, i + 1)$  and  $(0, k)$  for some  $1 \leq k \leq i + 1$ , then the common ancestry of  $(0, i + 2)$  and  $(0, i + 1)$  will be counted in the multiplicity of the mass at  $A_{i+1}$  in  $B_i$  and the updated version  $B_i^*$  will have nonzero multiplicity at  $A_{i+1}$  (so  $A_{i+1} = a_{1,i}^*$ ) and there will be no need to add a new point mass to create  $B_{i+1}$ . In addition, if this ancestor is also the last common ancestor of  $(0, i)$  and  $(0, i + 1)$  (so  $A_{i+1} = a_{1,i}$ ), the count for this ancestor cannot be 1 in  $B_i$ , so in this case we have  $A_{i+1} = a_{1,i} = a_{1,i}^*$  and there is no need to add a new point mass in  $B_{i+1}$ .

We now provide the law of this point measure process and its relationship to  $(A_i; i \geq 1)$ .

**THEOREM 2.2.** *The  $i$ th branch length is the smallest point mass in  $B_i$*

$$A_i = \mathfrak{s}(B_i) = \min\{n \geq 1 : B_i(\{n\}) \neq 0\}.$$

*In addition, the sequence of finite point measures  $(B_i; i \geq 0)$  is a Markov chain started at the null measure, such that for any finite point measure  $b = \sum_{n \geq 1} b_n \delta_n$ , with  $b_n \in \mathbb{N} \cup \{0\}$ , the law of  $B_{i+1}$  conditionally given  $B_i = b$ , is given by the*

following transition. Let  $a_1 := \mathfrak{s}(b)$ ,  $b^* := b - \delta_{a_1}$  and  $a_1^* := \mathfrak{s}(b^*)$ . Let  $(A, N)$  be distributed as  $(A_1, \zeta'_{A_1})$ . Then

$$B_{i+1} := \begin{cases} b^* + N\delta_A, & \text{if } A < a_1^* \text{ and } A \neq a_1, \\ b^*, & \text{otherwise.} \end{cases}$$

PROOF. Instead of the full information  $(D_1(n), n \geq 1)$ , this sequence starts with a single point measure

$$B_1 = D_1(A_1)\delta_{A_1},$$

and at each step it proceeds by changing the weights of the existing point masses and by adding at most one new point mass.

It is clear from the recursive definition of  $(B_i; i \geq 1)$  that for any  $i \geq 1$  if  $b_{n,i} \neq 0$ , then  $b_{n,i} = D_i(n)$ . We first show that for any  $i \geq 1$  we have  $a_{1,i} = A_i$  and  $b_{a_{1,i},i} = D_i(A_i)$ . If we show that  $b_{A_i,i} \neq 0$ , then, since all other nonzero weights in  $B_i$  satisfy  $b_{n,i} = D_i(n)$ , the definition of  $A_i$  will immediately imply that  $a_{1,i} = A_i$  and  $b_{a_{1,i},i} = D_i(A_i)$ . We do this by induction. The claim is clearly true for  $i = 1$ , so let us assume it is true for an arbitrary  $i \geq 1$ .

Consider what the transition rule for  $D_i$  tells us about the relationship between  $A_{i+1}$  and  $A_i$ . Recall for all  $n > A_i$  we have  $D_{i+1}(n) = D_i(n)$ ,  $D_{i+1}(A_i) = D_i(A_i) - 1$ , and for all  $n < A_i$  we have  $D_{i+1}(n) = \zeta'_n$  from r.v.s drawn independently of  $D_i$ . So,

$$\begin{aligned} A_{i+1} < A_i &\Leftrightarrow \exists n < A_i \text{ s.t. } \zeta'_n \neq 0, \\ A_{i+1} = A_i &\Leftrightarrow D_i(A_i) > 1 \text{ and } \zeta'_n = 0, \forall n < A_i, \\ A_{i+1} > A_i &\Leftrightarrow D_i(A_i) = 1 \text{ and } \zeta'_n = 0, \forall n < A_i. \end{aligned}$$

In the first case,  $A_{i+1} < A_i = a_{1,i} \leq a_{1,i}^*$ , so the point mass at  $A_{i+1}$  will be added to  $B_{i+1}$  and  $b_{A_{i+1},i+1} \neq 0$ . In the second case,  $b_{A_i,i+1} = D_{i+1}(A_i) = D_i(A_i) - 1 > 0$ , and since  $A_{i+1} = A_i$ , we have  $b_{A_{i+1},i+1} \neq 0$ . In the third case,  $b_{A_i,i+1} = D_{i+1}(A_i) = D_i(A_i) - 1 = 0$ , and also  $b_{a_{1,i},i} = D_i(A_i) = 1$  implies  $a_{1,i}^* = \min\{n > a_{1,i} : b_{n,i} \neq 0\} = \min\{n > A_i : b_{n,i} \neq 0\}$ . Note that for all  $n > A_i$  for which  $b_{n,i} \neq 0$ , we also have  $b_{n,i+1} = b_{n,i} \neq 0$ , and since  $b_{A_i,i+1} = 0$ , the smallest value of  $n$  with mass  $B_{i+1}(\{n\}) > 0$  before we potentially add a new mass is precisely at  $a_{1,i}^* = \min\{n > A_i : b_{n,i} \neq 0\}$ . In this case  $a_{1,i}^* > a_{1,i} = A_i$  so we have  $D_{i+1}(a_{1,i}^*) = D_i(a_{1,i}^*) \neq 0$ , so by definition of  $A_{i+1}$  we must have  $A_{i+1} \leq a_{1,i}^*$ . In case  $A_{i+1} < a_{1,i}^*$ , the point mass at  $A_{i+1}$  will be added to  $B_{i+1}$  and  $b_{A_{i+1},i+1} \neq 0$ . In case  $A_{i+1} = a_{1,i}^*$ , no new mass is added and the smallest of the nonzero masses in  $B_{i+1}$  is at  $a_{1,i}^* = A_{i+1}$ , as stated earlier, and again  $b_{A_{i+1},i+1} = b_{a_{1,i}^*,i+1} = b_{a_{1,i}^*,i} \neq 0$ . Hence, we have shown by induction that  $b_{A_i,i} \neq 0$  for all  $i \geq 1$ , so that  $a_{1,i} = A_i$  and  $b_{a_{1,i},i} = B(A_i, i)$ .

Now consider the transition rule for  $B_{i+1}$ , conditionally given  $B_i$ . For the already existing mass in  $B_i$  the changes in weights are given by the transition rule for  $D_i$  to be

$$\begin{aligned} \sum_{n \geq 1} D_{i+1}(n) \mathbf{1}_{\{b_{n,i} \neq 0\}} \delta_n &= (D_i(A_i) - 1) \delta_{A_i} + \sum_{n > A_i} D_i(n) \mathbf{1}_{\{b_{n,i} \neq 0\}} \delta_n \\ &= B_i - \delta_{a_{1,i}} = B_i^*. \end{aligned}$$

We have an addition of a new point mass iff  $A_{i+1} < a_{1,i}^*$ . Since  $a_{1,i}^* \geq a_{1,i} = A_i$ , this happens iff either:

- (i)  $A_{i+1} < A_i$ , or
- (ii)  $b_{a_{1,i},i} = 1$  and  $A_i < A_{i+1} < a_{1,i}^*$ .

The reason  $A_i = A_{i+1}$  is not included in (ii) is that if  $b_{a_{1,i},i} = D_i(A_i) = 1$  then  $D_{i+1}(A_i) = D_i(A_i) - 1 = 0$ . Let  $\{\zeta'_n\}_{n \geq 1}$  be a sequence of r.v.s drawn independently from  $D_i$ . Then, by the transition rule for  $D_i$ :

- (i) holds iff  $\exists n < A_i$  s.t.  $\zeta'_n \neq 0$ , then,  $A_{i+1} := \min\{n < A_i : \zeta'_n \neq 0\}$ ;
- (ii) holds iff  $b_{a_{1,i},i} = 1$ ,  $\forall n < A_i$   $\zeta'_n = 0$  and  $\exists n < a_{1,i}^*$  s.t.  $\zeta'_n \neq 0$ , then,  $A_{i+1} := \min\{A_i < n < a_{1,i}^* : \zeta'_n \neq 0\}$ .

In case (i) holds it is clear that the weight  $D_{i+1}(A_{i+1})$  is distributed as  $\zeta'_{A_1}$  conditional on  $A_1 < A_i = a_{1,i}^*$  and that this weight is independent of  $(B_j; 1 \leq j \leq i)$ . We next argue that this is also true in case (ii) holds. In this case  $A_i < A_{i+1} < a_{1,i}^*$  and since  $D_{i+1}(A_{i+1}) \neq 0$  we must also have  $D_i(A_{i+1}) \neq 0$  because the transition rule for  $D_{i+1}$  only allows zero entries to become nonzero for  $n < A_i$ . Moreover,  $D_i(A_{i+1}) = D_{i+1}(A_{i+1})$  because all entries for  $n > A_i$  remain unchanged from step  $i$  to  $i + 1$ . Hence, we must have  $D_n(A_{i+1}) = D_{i+1}(A_{i+1})$  for all  $k < n \leq i$  where

$$k := \max\{0 \leq j < i : A_j \geq a_{1,i}^*\}$$

with the convention that  $A_0 := +\infty$ .

We show that  $A_k \neq A_{i+1}$  if  $A_i < A_{i+1} < a_{1,i}^*$ . If we had  $A_k = A_{i+1}$ , then  $b_{A_k,k} = D_k(A_k) \neq 0$ . The transition rule for  $D_{k+1}$  implies  $D_{k+1}(A_k) = D_k(A_k) - 1$ . Since  $A_n < A_{i+1} = A_k$  for all  $k + 1 \leq n \leq i$ , iteratively applying the transition rule for  $D_{k+2}, \dots, D_{i+1}$  implies  $D_{k+2}(A_k) = D_{k+1}(A_k), \dots, D_{i+1}(A_k) = D_i(A_k)$ . Thus

$$D_k(A_k) - 1 = D_{k+1}(A_k) = \dots = D_i(A_k) = D_{i+1}(A_k) = D_{i+1}(A_{i+1}) \neq 0$$

and  $b_{A_k,k} = D_k(A_k) > 1$ . Then, by definition of the weights for  $B_{k+1}, \dots, B_{i+1}$ , we have that  $b_{A_k,k+1} = D_{k+1}(A_k) \neq 0$  and the same weight iteratively remains as  $b_{A_k,n} = D_n(A_k)$  for all  $k + 1 \leq n \leq i + 1$ . However,  $b_{A_{i+1},i} = b_{A_k,i} \neq 0$  contradicts our assumption that  $A_{i+1} < a_{1,i}^*$ .

Thus we must have  $A_k > A_{i+1}$ . Since  $D_i(A_k) \neq 0$  we also must have  $A_k \geq a_{1,i}^*$ . From the definition of weights for  $B_i$  and  $A_i < a_{1,i}^*$ , we have

$$b_{a_{1,i}^*,i} \neq 0 \quad \Rightarrow \quad \exists k' < i \text{ s.t. } a_{1,i}^* = A_{k'} \text{ and } A_n < A_{k'} \text{ for all } k' < n \leq i$$

since the only point mass in a step  $n$  not existing in the previous step must be placed at  $A_n$ . Now,  $A_{k'} = a_{1,i}^* > A_{i+1}$  and the fact that we defined  $k = \max\{0 \leq j < i : A_j \geq A_{i+1}\}$  implies that  $k = k'$ , hence,  $A_k = a_{1,i}^*$ .

Since  $A_k > A_{i+1}$ , we must have  $D_k(A_{i+1}) = 0$ . By the same argument as above of iteratively applying the transition rule for  $D_{k+2}, \dots, D_{i+1}$ , the weights at  $A_{i+1}$  satisfy

$$D_{k+1}(A_{i+1}) = \dots = D_i(A_{i+1}) = D_{i+1}(A_{i+1}) \neq 0.$$

Let us now consider the distribution of  $D_{i+1}(A_{i+1})$  conditional on the value of  $k$ . Since  $A_{i+1} < A_k$  and  $D_{k+1}(A_{i+1}) \neq 0$  by the transition rule for  $D_{k+1}$ , the value of the weight  $D_{k+1}(A_{i+1})$  is a r.v. distributed as  $\zeta'_{A_1}$  conditional on  $A_1 < A_k$  and it is drawn independently of  $(B_j; 1 \leq j \leq k)$ . So,  $D_{i+1}(A_{i+1}) = D_{k+1}(A_{i+1})$  is distributed as  $\zeta'_{A_1}$  conditional on  $A_1 < A_k = a_{1,i}^*$  and is independent of  $(B_j; 1 \leq j \leq k)$ . Since  $A_n < A_{i+1}$  for all  $k < n \leq i$ , the transition rule for  $D_{k+1}, \dots, D_i$  implies that all of the subsequently added point masses in  $B_{k+1}, \dots, B_i$  are independent of  $D_{k+1}(A_{i+1})$ , so the value of  $D_{i+1}(A_{i+1})$  is independent of  $(B_j; 1 \leq j \leq i)$ . Finally, integrating over  $k$ , we have that  $D_{i+1}(A_{i+1})$  is distributed as  $\zeta'_{A_1}$  conditional on  $A_1 < a_{1,i}^*$  and is independent of  $(B_j; 1 \leq j \leq i)$ .

We have now shown that when a new point mass is added at  $A_{i+1}$ , either  $A_{i+1} := \min\{n < A_i : \zeta'_n \neq 0\}$ , or  $b_{a_{1,i},i} = 1$  and  $A_{i+1} := \min\{A_i < n < a_{1,i}^* : \zeta'_n \neq 0\}$ , where the sequence  $\{\zeta'_n\}_{n \geq 1}$  are r.v.s drawn independently from  $(B_j; j \leq i)$ . In either case, the weight of the new point mass at  $A_{i+1}$  is distributed as  $\zeta'_{A_1}$  conditional on  $A_1 < a_{1,i}^*$ .

Let  $A := \min\{n \geq 1 : \zeta'_n \neq 0\}$  and  $N := \zeta'_A$ . Since  $\{\zeta'_n\}_{n \geq 1}$  are independent of  $(B_j; 1 \leq j \leq i)$ , so are  $A$  and  $N$ , and by Theorem 2.1,  $(A, N)$  is distributed as  $(A_1, \zeta'_{A_1})$ . Conditional on the value of  $B_i$ , we have that (i) holds iff  $A < A_i$ , while (ii) holds iff  $b_{a_{1,i},i} = 1$  and  $A_i < A < a_{1,i}^*$ . Putting (i) and (ii) together with the definition of  $a_{1,i}^*$  we have, conditionally on  $(B_j; 1 \leq j \leq i)$ , an addition of a new point mass iff  $A < a_{1,i}^*$  and  $A \neq a_{1,i}$ . Moreover, the weight of the newly added point mass is distributed as  $\zeta'_{A_1}$  conditional on  $A_1 < a_{1,i}^*$ , or equivalently it is distributed as  $N$  conditional on  $A < a_{1,i}^*$ . We also showed that given  $(B_j; 1 \leq j \leq i)$ , the point masses existing in  $B_i$  change in the next step to produce a re-weighted point mass measure equal to  $B_i - \delta_{a_{1,i}}$ . Altogether, given  $(B_j; 1 \leq j \leq i)$ , the next step of the sequence depends on  $B_i$  only, with the transition rule that if  $A < a_{1,i}^*$  and  $A \neq a_{1,i}$  in the next step we have  $B_{i+1} = B_i - \delta_{a_{1,i}} + N\delta_A$ , and otherwise in the next step we have  $B_{i+1} = B_i - \delta_{a_{1,i}}$ .  $\square$

In the course of the above proof, we have also shown that if  $t_1 := \min\{t \geq 1 : A_t > A_1\}$  and  $N_1 := \#\{1 \leq j < t_1 : A_j = A_1\}$ , then  $N_1 = D_1(A_1)$ , because in

order to add the first new point mass at a level greater than  $A_1$ , we must in the course of  $1 \leq j < t_1$  have exactly enough steps at which  $A_j = A_1$  that will exhaust all of the weight  $D_1(A_1)$  of the point mass at  $A_1$ .

Analogously for each  $i \geq 1$ , if we let

$$t_i := \min\{t \geq i : A_t > A_i\} \quad \text{and} \quad N_i := \#\{i \leq j < t_i : A_j = A_i\},$$

then  $D_i(A_i) = N_i$ . If, furthermore, for each  $i \geq 1$  and  $k \geq i$  we define

$$N_{ik} := \#\{k \leq j < t_i : A_j = A_i\},$$

then by a similar argument for  $k \geq i$  we have  $N_{ik} = D_k(A_i)$ . Note that  $N_{ii} \equiv N_i$ . Then, for the sequence of finite point measures we have that for all  $k \geq 1$ ,

$$(1) \quad B_k = \sum_{i \leq k} N_{ik} \delta_{A_i} \mathbf{1}_{\{A_i < A_{i'}, \forall i' < i : N_{i'k} \neq 0\}}.$$

It is easily checked that  $B_{k+1}$  correctly updates the weight of each existing point mass from  $B_k$ , and allows a new point mass to be added only if it is in a location smaller than all mass existing in  $B_k$  whose weights in  $B_{k+1}$  remain nonzero. We will see this formula again when we discuss the coalescent process in the continuous case.

REMARK 1. In the discrete case the coalescence times  $A_i$  take on integer values which may occur again after they have appeared for the last time  $t_i$  in a subtree. In the continuous case, analogously defined coalescence times have a law that is absolutely continuous w.r.t. Lebesgue measure, their values a.s. never occur again after they have appeared for the last time in a subtree, hence, in the continuous case there is no need to use separation times  $t_i$  in the definition of counters  $N_i$  and  $N_{ik}$ .

### 3. From discrete to continuous: The great-aunt measure.

3.1. *Definition of the great-aunt measure.* Now that we have a simple description of the coalescent point process with multiplicities for a BGW branching process, our goal is to do the same for the continuous-state branching process. In the discrete case we used the process  $(D_i; i \geq 1)$  to describe the number of surviving younger offshoots of ancestors of an individual as a sequence indexed by generations backward in time. Since in the CSB case the standing population is not discrete, we cannot use a process indexed by the standing population. Instead, we use a spine decomposition of the lineage of a surviving individual, which will record the level (i.e., generation in the discrete, and height in the continuous case) and the number of all offshoot subtrees in the individual's genealogy. We first provide the law of the spine decomposition of the first individual in the standing population in the BGW case and relate it to our previous results. At this point we would like to emphasize that results for the spine decomposition of BGW process are not new (see references at the end of the subsection), and that we make use of the spine

decomposition here only as a tool that will enable us to describe the analogue of the coalescent point process with multiplicities for the CSB process later.

We give some new definitions to describe the spine decomposition of  $(0, 1)$ , the first individual in generation 0, of a BGW process within its monotone planar embedding. For  $n \geq 1$ , we denote by  $\phi_n$  the set of great-aunts of  $(0, 1)$  at generation  $-n + 1$ , that is,

$$\phi_n := \{\text{daughters of } (-n, \alpha_1(n)) \text{ excluding } (-n + 1, \alpha_1(n - 1))\},$$

and by  $\varphi_n$  its cardinal,  $\varphi_n := \#\phi_n$ . In other words,  $\varphi_n$  is the number of offshoots of  $(-n, \alpha_1(n))$  not counting the lineage of  $(0, 1)$  itself. The set  $\phi_n$  can be divided into

$$\alpha_n := \{\text{sisters of } (-n + 1, \alpha_1(n - 1)) \text{ with labels } (-n + 1, k) \text{ s.t. } k < \alpha_1(n - 1)\}$$

the older offshoots of  $(-n, \alpha_1(n))$ , and

$$\rho_n := \{\text{sisters of } (-n + 1, \alpha_1(n - 1)) \text{ with labels } (-n + 1, k) \text{ s.t. } k > \alpha_1(n - 1)\}$$

the younger offshoots of  $(-n, \alpha_1(n))$ . We call the sequence  $(\rho_n; n \geq 1)$  the *great-aunt measure*. We let  $(Z_k^{\alpha_n}; 0 \leq k \leq n - 1)$  and  $(Z_k^{\rho_n}; 0 \leq k \leq n - 1)$  be the processes counting the descendants of those  $\alpha_n$  and  $\rho_n$  individuals, respectively, at successive generations  $-n + 1, \dots, 0$ .

An important observation now is that we do not need the whole infinite embedding of trees to make the previous definitions. The descending tree of  $(-h, \alpha_1(h))$  is a planar BGW tree conditioned to have alive individuals at generation  $h$ , with the lineage of individual  $(0, 1)$  being the left-most lineage with alive descendants at generation  $h$ . Consequently, provided that we only consider indices  $n \leq h$ , our definitions make sense for any conditioned BGW planar tree. A number of results have already been proved in [12] for spine decomposition of BGW process conditioned on survival at a given generation, and we make use of them here. The infinite embedding of trees that we introduced extends these results in a way, as we are considering an arbitrarily large standing population that may not [in (sub)critical case] descend from the same ancestor  $n$  generations back in the past. Considering trees whose roots are individuals  $(-n, \alpha_1(n)), (-n, \alpha_1(n) + 1), \dots$ , it is easy to see that they are independent and identically distributed as  $\mathcal{T}(n, 1)$ , the tree descending from  $(-n, \alpha_1(n))$ .

We recall some standard notation commonly used with linearly ordered planar trees. The *Ulam–Harris–Neveu* labeling of a planar tree assumes that each individual (vertex) of the tree is labeled by a finite word  $u$  of positive integers whose length  $|u|$  is the generation, or height, of the individual. A rooted, linearly ordered planar tree  $\mathcal{T}$  is a subset of the set  $\mathcal{U}$  of finite words of integers

$$\mathcal{U} = \bigcup_{n \geq 0} \mathbb{N}^n,$$

where  $\mathbb{N}^0$  is the empty set. More specifically, the root of  $\mathcal{T}$  is denoted  $\emptyset$  and her offspring are labeled  $1, 2, 3, \dots$  from left to right. The recursive rule is



that the offspring of any individual labeled  $u$  are labeled  $u1, u2, u3, \dots$  from left to right, where  $ui$  is the mere concatenation of the word  $u$  and the number  $i$ . The *depth-first search* is the lexicographical order associated with Ulam–Harris–Neveu labeling (see Figure 2, where the depth-first search gives the order  $\emptyset, 1, 11, 12, 2, 211, 212, 3, 31, 311, 32, \dots$ ).

Now fix  $h \in \mathbb{N}$  and assume that generation  $h$  is nonempty. Let  $x_h$  be the Ulam–Harris–Neveu label of the first individual in depth-first search with height  $h$ . We denote by  $(x_h \mid -n)$  the ancestor of  $x_h$  at generation  $h - n$ . Then for  $1 \leq n \leq h$ , we can define  $\alpha'_n$  and  $\rho'_n$  to be the number of daughters of  $(x_h \mid -n)$  ranked smaller and larger, respectively, than  $(x_h \mid -n + 1)$ , and define  $\varphi'_n := \alpha'_n + \rho'_n$ . We can also let  $Z^{\alpha'_n}$  and  $Z^{\rho'_n}$  be the processes counting the descendants of those  $\alpha'_n$  and  $\rho'_n$  individuals, respectively. From now on, we assume that the tree  $\mathcal{T}$  has the law of a planar BGW tree with offspring distributed as  $\xi$  and conditioned to have alive individuals at generation  $h$ . Then it is easily seen that  $(\alpha'_n, \rho'_n, Z^{\alpha'_n} \circ k_n, Z^{\rho'_n} \circ k_n; 1 \leq n \leq h)$  (where  $k_n$  means killing at time  $n$ ) has the same distribution as  $(\alpha_n, \rho_n, Z^{\alpha_n}, Z^{\rho_n}; 1 \leq n \leq h)$ , so from now on we remove primes. The following result provides the joint law of these random variables, which was already shown in [12], and the results below are just a restatement of Lemma 2.1 from [12]. Recall that  $p_n = \mathbb{P}(Z_n \neq 0 \mid Z_0 = 1)$ .

**PROPOSITION 3.1.** *Conditional on the values of  $(\alpha_n, \rho_n; n \geq 1)$ , the processes  $(Z^{\alpha_n}, Z^{\rho_n}; n \geq 1)$  are all independent,  $Z^{\rho_n}$  is a copy of  $Z$  started at  $\rho_n$  and killed at time  $n$  and  $Z^{\alpha_n}$  is a copy of  $Z$  started at  $\alpha_n$ , conditioned on being zero at time  $n - 1$ . In addition, the pairs  $(\alpha_n, \rho_n; n \geq 1)$  are independent and distributed as follows:*

$$\mathbb{P}(\alpha_n = j, \rho_n = k) = \mathbb{P}(\xi = k + 1) \frac{p_{n-1}}{p_n} (1 - p_{n-1})^j, \quad k \geq j \geq 0.$$

**REMARK 2.** Recall that  $D_1(n)$  is the number of sisters of  $(-n + 1, \mathbf{a}_1(n - 1))$  with alive descendants in the standing population. Thus, an immediate corollary of Proposition 3.1 is that the random variables  $(D_1(n); n \geq 1)$  are independent and that, conditionally on  $\rho_n$ ,  $D_1(n)$  is a binomial r.v. with parameters  $\rho_n$  and  $p_{n-1}$ . Since, according to Theorem 2.1,  $D_1(n)$  is distributed as  $\zeta'_n$ , we also have

$$\mathbb{E}(s^{\zeta'_n}) = \mathbb{E}((1 - p_{n-1} + p_{n-1}s)^{\rho_n})$$

as one can easily check.

Note that the last statement in Proposition 3.1 can be viewed as an inhomogeneous spine decomposition of the “*ascendance*” of the surviving particles in conditioned BGW trees. More standard spine decompositions are well known for the “*descendance*” of conditioned branching trees (see, e.g., [7, 9, 11, 15, 18, 20, 25]) (the idea of spine decompositions originating from [15]). To bridge the gap

between both aspects, notice that in the (sub)critical case,  $\mathbb{P}(\varphi_n + 1 = k) = \mathbb{P}(\xi = k)(1 - (1 - p_{n-1})^k)/p_n$  converges to  $k\mathbb{P}(\xi = k)/\mathbb{E}(\xi)$  as  $n \rightarrow \infty$ , which is the size-biased distribution of  $\xi$ . This distribution is known to be the law of offspring sizes on the spine when conditioning the tree on infinite survival.

3.2. *A random walk representation.* We next show how to recover any truncation  $(\rho_n; n \leq h)$  of the great-aunt measure from the paths of a conditioned, killed random walk. This will be particularly useful when we define the analogue of the great-aunt measure for continuous-state branching processes in the next subsection. Our result makes use of a well-known correspondence between a BGW tree and a downward-skip-free random walk introduced and studied in [3, 24].

Let us go back to the planar tree  $\mathcal{T}$ . We denote by  $v_n$  the word labeling the  $n$ th individual of the tree in the depth-first search. For any integers  $i > j$  and any finite word  $u$ , we say that  $ui$  is “younger” than  $uj$ . Also, we will write  $u < v$  if  $u$  is an ancestor of  $v$ , that is, there is a sequence  $w$  such that  $v = uw$  (in particular,  $u < u$ ). Last, for any individual  $u$  of the tree, we let  $r(u)$  denote the number of younger sisters of  $u$ , and for any integer  $n$ , we let  $W_n := 0$  if  $v_n = \emptyset$ , and if  $v_n$  is any individual of the tree different from the root, we let

$$W_n := \sum_{u < v_n} r(u).$$

The height  $H_n$ , or generation, of the individual visited at time  $n$  can be recovered from  $W$  as

$$(2) \quad H_n := |v_n| = \#\{1 \leq k \leq n : W_k = \min_{k \leq j \leq n} W_j\}.$$

See Figure 2 where  $(W_n; n = 1, \dots, 15)$  is given until visit of  $v_{15} = 3221$  whose height is  $H_{15} = 4$ .

In the case when  $\mathcal{T}$  is a BGW tree with offspring distributed as  $\xi$ , then it is known [3, 24] that the process  $(W_n; n \geq 1)$  is a random walk started at 0, killed upon hitting  $-1$ , with steps in  $\{-1, 0, 1, 2, \dots\}$  distributed as  $\xi - 1$ .

Fix  $h \in \mathbb{N}$  and set

$$\sigma_h := \min\{n \geq 1 : H_n = h\}.$$

Writing  $T_j$  for the first hitting time of  $j$ , in particular,  $T_{-1}$  for the first hitting time of  $-1$ , we get

$$(3) \quad \left\{ \max_{0 \leq j \leq T_{-1}} H_j < h \right\} = \{Z_h = 0\} = \{\sigma_h \geq T_{-1}\}.$$

Now, when  $\sigma_h < T_{-1}$ , we let  $I_{\sigma_h}$  denote the future infimum process of the random walk killed at  $\sigma_h$

$$I_{\sigma_h}^j := \min_{j \leq r \leq \sigma_h} W_r,$$

and we let  $t_h = 0, \dots, t_0 = \sigma_h$  denote the successive record times of  $I_{\sigma_h}$

$$t_k := \max\{j < t_{k-1} : W_j = I_{\sigma_h}^j\}, \quad 1 \leq k \leq h.$$

Also observe that by definition of  $\sigma_h$ , we must have  $t_1 = t_0 - 1$ . Last, we use the notation  $\Delta_j W := W_{j+1} - W_j$ . A straightforward consequence of [3, 24] is that when generation  $h$  is nonempty,  $t_k$  is the visit time of  $(x_h \mid -k)$ , that is, the unique integer  $n$  such that  $v_n = (x_h \mid -k)$  (where  $x_h$  is the first individual in depth-first search with height  $h$ ). Furthermore,

$$\varphi_k = \Delta_{t_k} W = W_{t_k+1} - W_{t_k}$$

and

$$\alpha_k = W_{t_k+1} - \min_{t_k+1 \leq j \leq t_0} W_j \quad \text{and} \quad \rho_k = \min_{t_k+1 \leq j \leq t_0} W_j - W_{t_k}.$$

In particular, we can check  $\alpha_1 = 0$ , and  $\varphi_k = \alpha_k + \rho_k$ . Now note that, by definition of  $t_h, \dots, t_0$ , we have  $I_{\sigma_h}^{t_k} = W_{t_k}$  and for all  $j : t_k < j < t_{k-1}$  we have  $I_{\sigma_h}^j = W_{t_{k-1}}$ , so

$$\alpha_k = W_{t_k+1} - I_{\sigma_h}^{t_k+1} = W_{t_k+1} - W_{t_{k-1}} \quad \text{and} \quad \rho_k = I_{\sigma_h}^{t_k+1} - W_{t_k} = W_{t_{k-1}} - W_{t_k}.$$

An illustration of these claims can be seen in Figure 2.

Last, observe that

$$\begin{aligned} \sum_{k=1}^h \rho_k f(k) &= \sum_{k=1}^h (I_{\sigma_h}^{t_k+1} - W_{t_k}) f(k) = \sum_{k=1}^h \sum_{j=0}^{\sigma_h-1} \mathbf{1}_{j=t_k} (I_{\sigma_h}^{j+1} - W_j) f(h - H_j) \\ &= \sum_{j=0}^{\sigma_h-1} \Delta_j I_{\sigma_h}^j f(h - H_j) \end{aligned}$$

since, if  $j = t_k$  for some  $k$ , then

$$\Delta_j I_{\sigma_h}^j = \Delta_{t_k} I_{\sigma_h}^{t_k} = I_{\sigma_h}^{t_k+1} - I_{\sigma_h}^{t_k} = I_{\sigma_h}^{t_k+1} - W_{t_k},$$

otherwise,  $t_k < j < t_{k-1}$  for some  $k$ , and

$$\Delta_j I_{\sigma_h}^j = I_{\sigma_h}^{j+1} - I_{\sigma_h}^j = W_{t_{k-1}} - W_{t_{k-1}} = 0.$$

This is recorded in the following statement, which we will use later as a distributional equality for a random walk  $W$  conditioned on  $\max_{1 \leq j \leq T-1} H_j \geq h$ .

**PROPOSITION 3.2.** *For any  $f : \mathbb{N} \rightarrow \mathbb{R}_+$  with bounded support,  $h > \text{Supp}(f)$ , if  $\max_{1 \leq j \leq T-1} H_j \geq h$ , then*

$$\langle \rho, f \rangle := \sum_{k \geq 1} \rho_k f(k) = \sum_{j=0}^{\sigma_h-1} \Delta_j I_{\sigma_h}^j f(h - H_j),$$

where  $\Delta_j I_{\sigma_h}^j = I_{\sigma_h}^{j+1} - I_{\sigma_h}^j$ .

3.3. *A continuous version of the great-aunt measure.* In Section 2, we gave a consistent way of embedding trees with an arbitrary size of the standing population, each descending from an arbitrarily old founding ancestor, so that the descending subtree of each vertex is a BGW tree. The natural analog of this presentation is the flow of subordinators introduced by Bertoin and Le Gall in [4]. Because the Poissonian construction of this flow displayed in [5], Section 2, only holds for critical CSB processes and without Gaussian component, and because it is rather awkward to use in order to handle the questions we address here, we will now define an analogue of the great-aunt measure, using the genealogy of continuous-state branching processes introduced in [24] and further investigated in [8].

We start with a Lévy process  $X$  with no negative jumps and Laplace exponent  $\psi$  started at  $x > 0$ , such that  $\psi'(0+) \geq 0$ , so that  $X$  hits 0 a.s. As specified in [21, 23], the path of  $X$  killed upon reaching 0 can be seen as the (jumping) contour process of a continuous tree whose ancestor is the interval  $(0, x)$ . For example, the excursions of  $X$  above its past infimum  $I$  are the contour processes of the offspring subtrees of the ancestor. Almost surely for all  $t$ , it is possible to define the *height* (i.e., generation)  $H_t$  of the point of the tree visited at time  $t$ , as

$$(4) \quad H_t := \lim_{k \rightarrow \infty} \frac{1}{\varepsilon_k} \int_0^t \mathbf{1}_{\{X_s < I_t^s + \varepsilon_k\}} ds < \infty,$$

where  $(\varepsilon_k)$  is some specified vanishing positive sequence, and

$$I_t^s := \inf_{s \leq r \leq t} X_r.$$

With this definition, one can recover the population size at generation  $a$  as the density  $Z_a$  of the occupation measure of  $H$  at  $a$ , that is, the (total) local time of  $H$  at level  $a$ . It is proved in [8, 24] that this local time exists a.s. for all  $a$  and that  $(Z_a; a \geq 0)$  is a continuous-state branching process with branching mechanism  $\psi$ .

From now on, we will deal with a general branching mechanism  $\psi$  characterized by its Lévy–Khinchin representation

$$\psi(\lambda) = a\lambda + \beta\lambda^2 + \int_{(0, \infty)} \Lambda(dr)(e^{-\lambda r} - 1 + \lambda r \mathbf{1}_{\{r < 1\}}),$$

where  $\beta > 0$  is the *Gaussian component* and  $\Lambda$  is a positive measure on  $(0, \infty)$  such that  $\int_{(0, \infty)} (1 \wedge r^2) \Lambda(dr) < \infty$ , called the *Lévy measure*. We will denote by  $X$  a Lévy process with Laplace exponent  $\psi$  (started at 0 unless otherwise stated), and by  $Z$  a continuous-state branching process, or CSB process, with branching mechanism  $\psi$ .

We (only) make the following two assumptions. First,

$$\sup\{\lambda : \psi(\lambda) \leq 0\} =: \eta < \infty,$$

so that  $X$  is not a subordinator (i.e., it is not a.s. nondecreasing). Second,

$$\int_{[\eta+1, \infty)} \frac{du}{\psi(u)} < \infty,$$

so that  $Z$  either is absorbed at 0 in finite time or goes to  $\infty$ , and  $H$  has *a.s. continuous sample paths* [24]. This also forces  $\int_{(0,1)} r \Lambda(dr)$  to be infinite, so that the paths of  $X$  have *infinite variation*. We further set

$$\phi(\lambda) := \int_{[\lambda, \infty)} \frac{du}{\psi(u)}, \quad \lambda > \eta,$$

and  $v$  the inverse of  $\phi$  on  $(\eta, +\infty)$

$$v(x) := \phi^{-1}(x), \quad x \in (0, \infty).$$

Notice that  $v$  is nonincreasing and has limit  $\eta$  at  $+\infty$ . It is well known (e.g., [21]) that if  $Z$  is started at  $x$ , then it is absorbed at 0 before time  $t$  with probability  $e^{-xv(t)}$ . Also, if  $N$  denotes the excursion measure of  $X - I$  away from 0 under  $\mathbb{P}$  (normalized so that  $-I$  is the local time), then by [8], Corollary 1.4.2,

$$N(\sup H > a) = v(a), \quad a > 0.$$

Here, we want to allow the heights of the tree to take negative values. To do this, we start with a measure which embodies the mass distribution broken down on heights, of the population whose descendances have not yet been visited, in the same vein as in [8, 24], but with negative heights. In the usual setting, the mass distribution  $\rho_t$  of the population whose descendances have not yet been visited by the contour process  $X$  by time  $t$ , is defined by

$$\langle \rho_t, f \rangle = \int_{[0,t]} d_s I_t^s f(H_s)$$

for any nonnegative function  $f$ , where on the right-hand side we mean integrating the function  $s \mapsto f(H_s)$  with respect to the Stieltjes measure associated with the function  $s \mapsto I_t^s$ . Then  $\rho_t([a, b])$  is the mass of the tree between heights  $a$  and  $b$  whose descendants have not yet been visited by the contour process.

Here, we will start with a random positive measure  $\rho^0$  on  $[0, +\infty)$ , with the interpretation that  $\rho^0([a, b])$  is the mass of the tree between heights  $-b$  and  $-a$  whose descendants have not yet been visited by the contour process. This measure is the exact analogue of the great-aunt measure of the previous subsection.

DEFINITION 3.3. For every  $t > 0$ , set  $\pi^{(t)}(dr) := p(t, r) dr$ , where

$$p(t, r) := e^{v(t)r} \int_{(r, \infty)} e^{-v(t)z} \Lambda(dz), \quad r > 0.$$

We define  $\rho^0$  in law by

$$\rho^0(dx) := \beta dx + \sum_{t: \Delta_t > 0} \Delta_t \delta_t(dx),$$

where  $\delta$  denotes a Dirac measure, and  $(t, \Delta_t)$  is a Poisson point measure with intensity measure  $dt \pi^{(t)}(dr)$ .

3.4. *Convergence of the great-aunt measure.* We can now prove a theorem yielding two justifications for the definition of the measure  $\rho^0$ . First, we show that  $\rho^0(dx)$  as defined above is indeed the mass, at height  $h - x$ , of the part of the tree whose descendants have not yet been visited, either by a long-lived contour process ( $h \rightarrow \infty$ ) or under the measure  $N(\cdot \mid \sup H > h)$ . Second, we show the convergence of the appropriately re-scaled discrete great-aunt measures to the measure  $\rho^0$ , as the BGW processes approach the CSB process  $Z$  with branching mechanism  $\psi$ . In the next section we will show how  $\rho^0$  allows us to define the coalescent point process with multiplicities for the CSB process, and help us establish convergence from the appropriately re-scaled point process  $(B_i; i \geq 1)$ .

We assume there exists a sequence  $(\gamma_p; p \geq 1)$ ,  $\gamma_p \rightarrow \infty$  as  $p \rightarrow \infty$ , and a sequence of random variables  $(\xi_p; p \geq 1)$ , such that, if  $W^{(p)}$  denotes the random walk with steps distributed as  $\xi_p - 1$ , the random variables  $(p^{-1}W_{[p\gamma_p]}^{(p)})$  converge in law to  $X_1$ , where  $X$  denotes the Lévy process with Laplace exponent  $\psi$ . Then it is known ([14], Theorems 3.1 and 3.4) that if  $Z^{(p)}$  denotes the BGW process started at  $[px]$  with offspring size distributed as  $\xi_p$ , then the re-scaled processes  $(p^{-1}Z_{[\gamma_p t]}^{(p)}; t \geq 0)$  converge weakly in law in Skorokhod space to the CSB process  $Z$  with branching mechanism  $\psi$ , started at  $x$ .

In the following statement, we denote by  $\rho^{(p)}$  the great-aunt measure associated to the offspring size  $\xi_p$ . We have to make the following additional assumptions: if  $f^{(p)}$  denotes the p.g.f. of  $\xi_p$  and  $f_n^{(p)}$  its  $n$ th iterate, then for each  $\delta > 0$ ,

$$\liminf_{p \rightarrow \infty} f_{[\delta\gamma_p]}^{(p)}(0)^p > 0.$$

Convergence results in (iii) below rely heavily on results already established by Duquesne and Le Gall in [8] on convergence of appropriately re-scaled random walks  $W^{(p)}$  and their height processes  $H^{(p)}$  to the Lévy process  $X$  and its height process  $H$ . In particular, technical conditions such as the one above are justified in [8], Section 2.3.

**THEOREM 3.4.** *Let  $\sigma_h$  denote the first time that  $H$  hits  $h > 0$ .*

(i) *For any  $h > 0$  and for any nonnegative Borel function  $f$  vanishing outside  $[0, h]$ , the random variable  $\int_{[0, \sigma_h]} d_s I_{\sigma_h}^s f(h - H_s)$ , under  $N(\cdot \mid \sup H > h)$ , has the same law as  $\langle \rho^0, f \rangle$ ;*

(ii) *For any nonnegative Borel function  $f$  with compact support, as  $h \rightarrow \infty$  the random variables  $\int_{[0, \sigma_h]} d_s I_{\sigma_h}^s f(h - H_s)$  under  $\mathbb{P}(\cdot \mid \sigma_h < \infty)$  converge in distribution to  $\langle \rho^0, f \rangle$ ;*

(iii) *For any nonnegative Borel function  $f$  with compact support and sequence of nonnegative continuous functions  $\{f_p\}$  such that  $f_{p, \gamma_p}(x_p) := f_p(\gamma_p^{-1}x_p) \rightarrow f(x)$  whenever  $\gamma_p^{-1}x_p \rightarrow x$ , the random variables  $p^{-1}\langle \rho^{(p)}, f_{p, \gamma_p} \rangle$  converge in distribution to  $\langle \rho^0, f \rangle$  as  $p \rightarrow \infty$ .*

PROOF. Let us prove (i). It is known [8] that a.s. for all  $t$  the inverse of the local time on  $[0, t]$  of the set of increase times of  $(I_t^s; s \in [0, t])$  has drift coefficient  $\beta$ , so that

$$\int_{[0,t]} d_s I_t^s f(H_s) = \beta \int_0^{H_t} f(x) dx + \sum_{s \in [0,t]} f(H_s) \Delta_s I_t^s,$$

where the sum is taken over all times  $s$  at which  $(I_t^s; s \in [0, t])$  has a jump, whose size is then denoted  $\Delta_s I_t^s$ . This set of times will be denoted  $\mathcal{J}_h$  when  $t = \sigma_h$ . As a consequence, it only remains to prove that the random point measure  $M_h$  on  $(0, h) \times (0, \infty)$  defined by

$$M_h := \sum_{s \in \mathcal{J}_h} \delta(H_s, \Delta_s I_{\sigma_h}^s),$$

where the sum is zero when  $\mathcal{J}_h$  is empty ( $\sigma_h = \infty$ ), is an inhomogeneous Poisson point measure with the correct intensity measure. More precisely, we are going to check that for any nonnegative two-variable Borel function  $f$

$$\begin{aligned} N \left( \int_{(0,h) \times (0,\infty)} M_h(dt dr) f(h-t, r) \mid \sup H > h \right) \\ = \int_0^h dt \int_0^\infty dr p(t, r) f(t, r), \end{aligned}$$

that is,

$$N \int_{(0,h) \times (0,\infty)} M_h(dt dr) f(h-t, r) = v(h) \int_0^h dt \int_0^\infty dr p(t, r) f(t, r).$$

First notice that

$$\begin{aligned} N \int_{(0,h) \times (0,\infty)} M_h(dt dr) f(h-t, r) \\ = N \sum_{s: \Delta X_s > 0} \mathbf{1}_{s < \sigma_h} f(h - H_s, \Delta X_s + K'_s) \mathbf{1}_{\Delta X_s > -K'_s}, \end{aligned}$$

where  $K'_s$  is the global infimum of the shifted path  $X'$

$$X'_u := X_{s+u} - X_s, \quad 0 \leq u \leq \sigma_h - s.$$

But on  $\{s < \sigma_h\}$ ,  $\sigma_h - s$  is also  $\sigma'_{h-H_s}$  (with obvious notation), so by predictable projection and by the compensation formula applied to the Poisson point process of jumps, we get

$$N \int_{(0,h) \times (0,\infty)} M_h(dt dr) f(h-t, r) = N \int_0^{\sigma_h} ds \int_{(0,\infty)} \Lambda(dz) G_f(z, H_s),$$

where (with the notation  $I$  for the current infimum)

$$G_f(z, t) = \mathbb{E}_0[f(h-t, z + I_{\sigma_{h-t}}) \mathbf{1}_{-I_{\sigma_{h-t}} < z}].$$

Now  $-I_{\sigma_{h-t}}$  is the local time at the first excursion of  $X - I$  with height larger than  $h - t$  so it is exponentially distributed with parameter  $N(\sup H > h - t) = v(h - t)$ . As a consequence,

$$G_f(z, t) = v(h - t) \int_0^z dr f(h - t, r) e^{-(z-r)v(h-t)},$$

which yields

$$N \int_{(0,h) \times (0,\infty)} M_h(dt dr) f(h - t, r) = N \int_0^{\sigma_h} ds F_f(h - H_s)$$

with

$$\begin{aligned} F_f(t) &= \int_{(0,\infty)} \Lambda(dz) G_f(z, h - t) \\ &= v(t) \int_{(0,\infty)} \Lambda(dz) \int_0^z dr f(t, r) e^{-(z-r)v(t)} \\ &= v(t) \int_0^\infty dr f(t, r) e^{rv(t)} \int_{(r,\infty)} \Lambda(dz) e^{-zv(t)} \\ &= v(t) \int_0^\infty dr p(t, r) f(t, r). \end{aligned}$$

So we only have to verify that for any nonnegative Borel function  $g$ ,

$$N \int_0^{\sigma_h} ds g(h - H_s) = \int_0^h dt \frac{v(h)}{v(t)} g(t).$$

Due to results in [24], there indeed is a jointly measurable process  $(Z(a, t); a, t \geq 0)$  such that a.s. for all  $a$ ,

$$\int_0^a ds g(H_s) = \int_0^\infty dt Z(a, t) g(t).$$

In particular,

$$N \int_0^{\sigma_h} ds g(h - H_s) = \int_0^h dt N(Z(\sigma_h, t)) g(h - t),$$

so we just need to check that  $N(Z(\sigma_h, t)) = v(h)/v(h - t)$ , that is,

$$N(Z(\sigma_h, t) \mid \sup H > h) = \frac{1}{v(h - t)}.$$

But, conditional on  $\sup H > h$ ,  $\sigma_t < \infty$  and  $Z(\sigma_h, t)$  is the local time of  $H$  at level  $t$  between  $\sigma_t$  and  $\sigma_h$ , which is exponential with parameter  $N(\sup H > h - t) = v(h - t)$ , hence, the claimed expectation.

We proceed with (ii), which readily follows from (i). Indeed, for any  $h$  larger than  $\sup\{x : f(x) \neq 0\}$ ,

$$\int_{[0,\sigma_h]} d_s I_{\sigma_h}^s f(h - H_s) = \int_{[0,\sigma_h]} d_s I_{\sigma_h}^{s'} f(h - H_s'),$$



where primes indicate that the future infimum and the height process are taken w.r.t. the process  $X'$  with law  $N(\cdot \mid \sup H > h)$ , defined as

$$X'_t = X_{\rho_h+t} - I_{\sigma_h}, \quad 0 \leq t \leq \tau_h - \rho_h,$$

where  $\rho_h$  is the unique time  $s \leq \sigma_h$  when  $X_s = I_{\sigma_h}$  and  $\tau_h$  is the first time  $t \geq \sigma_h$  when  $X_t = I_{\sigma_h}$ .

We end the proof with (iii). Thanks to Proposition 3.2, we know that  $\langle \rho^{(p)}, f_{p,\gamma_p} \rangle$  has the same law as

$$\sum_{j=0}^{[\sigma_{\gamma_p h}^p]} \Delta I_{\sigma_{\gamma_p h}^p}^{(p)j} f_{p,\gamma_p}(\gamma_p h - H_j^{(p)})$$

conditional on  $\max_{1 \leq j \leq T_{-1}^{(p)}} H_j^{(p)} \geq \gamma_p h$ , where  $H^{(p)}$  is the height process associated with  $W^{(p)}$ ,  $T_{-1}^{(p)}$  is the first hitting time of  $-1$  by  $W^{(p)}$ ,  $\sigma_{\gamma_p h}^p$  denotes the first hitting time of  $\gamma_p h$  by  $H^{(p)}$  and  $I^{(p)}$  denotes the future infimum process of  $W^{(p)}$  stopped at time  $\sigma_{\gamma_p h}^p$ . If we can prove convergence of  $(p^{-1}W_{[p\gamma_p \wedge \sigma_{\gamma_p h}^p]}^{(p)}, \gamma_p^{-1}H_{[p\gamma_p \wedge \sigma_{\gamma_p h}^p]}^{(p)})$  under this conditional law to  $(X_{\cdot \wedge \sigma_h}, H_{\cdot \wedge \sigma_h})$  under the measure  $N(\cdot \mid \sup H \geq h)$ , then by the generalized continuous mapping theorem (e.g., [16], Theorem 4.27) we will get the convergence of

$$\sum_{j=0}^{[\sigma_{\gamma_p h}^p]} \frac{\Delta I_{\sigma_{\gamma_p h}^p}^{(p)j}}{p} f_p\left(h - \frac{H_j^{(p)}}{\gamma_p}\right),$$

which has the law of  $p^{-1}\langle \rho^{(p)}, f_{p,\gamma_p} \rangle$ , to  $\int_{[0,\sigma_h]} ds I_{\sigma_h}^s f(h - H_s)$  under  $N(\cdot \mid \sup H > h)$  which has the law of  $\langle \rho^0, f \rangle$  thanks to (i).

For (sub)critical  $\xi_p$  the convergence of this pair in distribution on the Skorokhod space of cadlag real-valued paths is a direct consequence of the results Corollary 2.5.1 and Proposition 2.5.2 already shown in [8], Section 2.5. To verify that the same holds for supercritical  $\xi_p$  as well, we note that the assumption of (sub)criticality (hypothesis (H2) in the notation of [8]) is not crucial in any of the steps of their proof, since the obtained convergence relies essentially only on the assumption that  $p^{-1}W_{[p\gamma_p \cdot]}^{(p)}$  converges to  $X$  (see the comment in the proof of Theorem 2.2.1 in [8] that any use of subcriticality in their proof can be replaced by using weak convergence of random walks). Of course, in the supercritical case, the height process  $H$  may drift off to infinity corresponding to the event that the CSB process  $Z$  survives forever, in which case it will code only an incomplete part of the genealogy of the first lineage which survives forever. However, for our purposes we only need to consider the genealogy of the first lineage that survives until time  $h$ , so this will not be an impediment for our considerations.

For supercritical  $\xi_p$  we still have the convergence of the pair

$$(5) \quad (p^{-1}W_{[p\gamma_p \cdot]}^{(p)}, \gamma_p^{-1}H_{[p\gamma_p \cdot]}^{(p)}) \xrightarrow[p \rightarrow \infty]{d} (X \cdot, H)$$

in distribution on the Skorokhod space of cadlag real-valued paths as in Theorem 2.3.1 and the first part of Corollary 2.5.1 in [8], Section 2.5. We now follow the same reasoning as in [8], Proposition 2.5.2. Let  $G_{\gamma_p h}^p = \sup\{s \leq \sigma_{\gamma_p h}^p : H_s^{(p)} = 0\}$ . If we think of  $H^{(p)}$  as the height process for a sequence of independent BGW trees with offspring distribution  $\xi_p$ , then  $G_{\gamma_p h}^p$  is the initial point of the first BGW tree in this sequence which reaches a height  $\gamma_p h$ . Let  $(\tilde{W}^{(p)}, \tilde{H}^{(p)})$  denote the process obtained by conditioning  $(W^{(p)}, H^{(p)})$  on  $\max_{1 \leq j \leq T_{-1}^{(p)}} H_j^{(p)} \geq \gamma_p h$ . Then,

$$\begin{aligned} & (p^{-1}\tilde{W}_{[p\gamma_p \cdot \wedge \sigma_{\gamma_p h}^p]}^{(p)}, \gamma_p^{-1}\tilde{H}_{[p\gamma_p \cdot \wedge \sigma_{\gamma_p h}^p]}^{(p)}) \\ & \stackrel{d}{=} (p^{-1}W_{[p\gamma_p(G_{\gamma_p h}^p + \cdot) \wedge \sigma_{\gamma_p h}^p]}^{(p)}, \gamma_p^{-1}H_{[p\gamma_p(G_{\gamma_p h}^p + \cdot) \wedge \sigma_{\gamma_p h}^p]}^{(p)}). \end{aligned}$$

Let  $G_h = \sup\{s \leq \sigma_h : H_s = 0\}$ , and let  $(\tilde{X}, \tilde{H})$  denote the process obtained by conditioning  $(X, H)$  on  $\sup H \geq h$ , then

$$(\tilde{X}_{\cdot \wedge \sigma_h}, \tilde{H}_{\cdot \wedge \sigma_h}) \stackrel{d}{=} (X_{(G_h + \cdot) \wedge \sigma_h}, H_{(G_h + \cdot) \wedge \sigma_h}).$$

If we use the Skorokhod representation theorem to assume that the convergence (5) holds a.s., the same arguments as in proof Proposition 2.5.2 of [8] imply that  $\sigma_{\gamma_p h}^p$  converges a.s. to  $\sigma_h$  and  $G_{\gamma_p h}^p$  converges a.s. to  $G_h$ , and hence,  $(p^{-1}W_{[p\gamma_p(G_{\gamma_p h}^p + \cdot) \wedge \sigma_{\gamma_p h}^p]}^{(p)}, \gamma_p^{-1}H_{[p\gamma_p(G_{\gamma_p h}^p + \cdot) \wedge \sigma_{\gamma_p h}^p]}^{(p)})$  converges a.s. to  $(X_{(G_h + \cdot) \wedge \sigma_h}, H_{(G_h + \cdot) \wedge \sigma_h})$ , proving that

$$(p^{-1}\tilde{W}_{[p\gamma_p \cdot]}^{(p)}, \gamma_p^{-1}\tilde{H}_{[p\gamma_p \cdot]}^{(p)}) \xrightarrow[p \rightarrow \infty]{d} (\tilde{X} \cdot, \tilde{H} \cdot)$$

in distribution on the Skorokhod space of cadlag real-valued paths. From this the desired convergence in distribution of  $p^{-1} \langle \rho^{(p)}, f_{p, \gamma_p} \rangle$  conditional on  $\max_{1 \leq j \leq T_{-1}^{(p)}} H_j^{(p)} \geq \gamma_p h$  to  $\langle \rho^0, f \rangle$  under the measure  $N(\cdot \mid \sup H \geq h)$  follows.  $\square$

#### 4. The coalescent point process in the continuous case.

4.1. *Definition of the genealogy.* We now define the analogue of a coalescent point process with multiplicities for the genealogy (other than the immediately recent) of an arbitrarily large standing population of a CSB process  $Z$ . We do so by first constructing the height process,  $H^*$ , for a planar embedding of CSB trees of

arbitrary size descending from an arbitrarily old ancestor using the continuous version of the great-aunt measure,  $\rho^0$ . From  $H^*$  we define coalescence times of two masses from the standing population in the usual way, that is, from the maximal depths of the trajectory of  $H^*$ .

We construct the height  $H_t^*$  of the individual visited at time  $t$  from: the height  $H_t$  defined in (4) representing the height at which that individual occurs in an excursion of  $X$  above its infimum; plus the height at which this excursion branches off the unexplored part of the tree. More specifically, let

$$Y_x^0 := \rho^0([0, x]), \quad x \geq 0,$$

so that

$$Y_x^0 = \beta x + \sum_{t \leq x} \Delta_t,$$

where  $(t, \Delta_t)$  is a Poisson point measure with intensity given by  $dt \pi^{(t)}(dr)$  from Definition 3.3. Next, let  $L^0$  denote the right-inverse of  $Y^0$ ,

$$L^0(t) := \inf\{a : Y_a^0 > t\}, \quad t \geq 0.$$

Then define

$$H_t^* := H_t - L^0(-I_t),$$

where we remember that  $I_t = \inf_{0 \leq s \leq t} X_s$ .

This gives a spine decomposition of the genealogy of the continuous-state branching process associated with  $H^*$  in the following sense. For the individual visited by the traversal process at time  $t$ , the level (measured back into the past, with the present having level 0) on the infinite spine at which the subtree containing this individual branches off is  $-L^0(-I_t)$ , and the relative height of this individual within this subtree is  $H_t$ .

As in the discrete case, we want to display the law of the coalescence time between successive individuals at generation 0. In this setting, this corresponds to the maximum depth below 0 of the height process, between successive visits of 0. Actually, any point at height 0 is a point of accumulation of other points at height 0, so we discretize the population at height 0 as follows. We consider all points at height 0 such that the height process between any two of them goes below  $-\varepsilon$ , for some fixed  $\varepsilon > 0$ , namely, we set  $T_0 := 0$ , and for any  $i \geq 1$ ,

$$S_i := \inf\{t \geq T_{i-1} : H_t^* = -\varepsilon\} \quad \text{and} \quad T_i := \inf\{t \geq S_i : H_t^* = 0\}.$$

Then the coalescence times  $A_1^\varepsilon, A_2^\varepsilon, \dots$  of the  $\varepsilon$ -discretized population are defined as

$$(6) \quad A_i^\varepsilon := -\inf\{H_t^* : T_{i-1} \leq t \leq T_i\}.$$

As in the discrete case, one of the main difficulties lies in the fact that the same value of  $A_i^\varepsilon$  can be repeated several times. So we define

$$(7) \quad N_i^\varepsilon := \#\{j \geq i : A_j^\varepsilon = A_i^\varepsilon\}.$$

4.2. *The supercritical case.* Actually, the previous definition of genealogy only holds for the subcritical and critical cases, and a modification needs to be made in the supercritical case due to possible appearances of branches with infinite survival times into the future. What we need to do in the supercritical case is construct a height process  $H^*$  that corresponds to a CSB tree whose infinite branches have been truncated, much as in the last chapter of [21]. One would be naturally led to consider the height process of a tree truncated at some finite height. However, the distribution of such an object is far more complicated than to truncate (actually, reflect) the associated Lévy process at some finite level. On the other hand, the genealogy coded by a Lévy process reflected below some fixed level is not easy to read since it will have incomplete subtrees at different heights. To comply with this difficulty, we will construct a consistent family of Lévy processes  $X^\kappa$  reflected below level  $\kappa$ , so that if  $\kappa' > \kappa$ ,  $X^\kappa$  can be obtained from  $X^{\kappa'}$  by excising the subpaths above  $\kappa$ . The genealogy coded by the projective limit of this family is the supercritical Lévy tree, as is shown in [21] in the case with finite variation. We first give the details of this construction in the discrete case and then sketch its definition in the CSB case. Let  $\mathcal{T}$  be a planar, discrete, possibly infinite, tree in the sense that it can have infinite height but all finite breadths. Then  $\mathcal{T}$  still admits a Ulam–Harris–Neveu labeling and, as in Section 3.2, we can define  $\hat{W}(\emptyset) = 0$  and

$$\hat{W}(v) := \sum_{u < v} r(u), \quad u \neq \emptyset,$$

where  $r(u)$  denotes the number of younger sisters of  $u$ . Now for any  $\kappa \in \mathbb{N}$ , let  $\mathcal{T}^\kappa$  denote the graph obtained from  $\mathcal{T}$  by deleting all vertices  $v$  such that  $\hat{W}(v) > \kappa$ . It is then easy to see that  $\mathcal{T}^\kappa$  is still a tree, and that if  $v_n^\kappa$  denotes the  $n$ th vertex of  $\mathcal{T}^\kappa$  in the lexicographical order and  $W_n^\kappa := \hat{W}(v_n^\kappa)$ , then for any  $\kappa' > \kappa$ , the path of  $W^\kappa$  can be obtained from the path of  $W^{\kappa'}$  by excising the subpaths above  $\kappa$ . In other words,

$$W^\kappa = C_\kappa(W^{\kappa'}),$$

where for any path  $\epsilon$ , the functional  $C_\kappa$  erasing subpaths above  $\kappa$  can be defined as follows. Let  $A^\kappa$  denote the additive functional

$$A_n^\kappa(\epsilon) := \sum_{k=0}^n \mathbf{1}_{\{\epsilon_k \leq \kappa\}}, \quad n \in \mathbb{N},$$

and  $a^\kappa$  its right inverse

$$a_k^\kappa(\epsilon) := \min\{n : A_n^\kappa(\epsilon) > k\}, \quad k \in \mathbb{N}.$$

Then  $C_\kappa(\epsilon) := \epsilon \circ a^\kappa(\epsilon)$ .

Further, it can be shown that  $W^\kappa$  has the law of the random walk with steps distributed as  $\xi - 1$ , reflected below  $\kappa$  and killed upon hitting  $-1$ . Since these two properties (consistency by truncation and marginal distribution) characterize the

family of reflected, killed random walks ( $W^\kappa; \kappa \in \mathbb{N}$ ), its continuous analogue has the following natural definition.

For any  $\kappa > 0$  let  $X^\kappa$  denote a Lévy process without negative jumps and Laplace exponent  $\psi$  reflected below  $\kappa$  and killed upon reaching 0. The reflection can be properly defined as follows. Start with the path of a Lévy process  $X$  (with the same Laplace exponent), set  $S_t := \sup_{0 \leq s \leq t} X_s$  and define  $X_t^\kappa := X_t$  if at time  $t$ ,  $X$  has not yet hit  $(\kappa, +\infty)$ , and  $X_t^\kappa := \kappa + X_t - S_t$  otherwise (and kill  $X^\kappa$  when it hits 0). The same definition could be done by concatenating excursions of  $X$  away from  $(\kappa, +\infty)$  thanks to Itô's synthesis theorem.

Now, in the continuous setting, a similar definition of  $C_\kappa$  can be done by adapting the additive functional  $A^\kappa$  into

$$A_t^\kappa(\epsilon) := \int_0^t ds \mathbf{1}_{\{\epsilon_s \leq \kappa\}}, \quad t \geq 0.$$

It is easily seen that for any  $\kappa' > \kappa$ ,

$$X^\kappa \stackrel{d}{=} C_\kappa(X^{\kappa'}).$$

Kolmogorov's extension theorem then ensures the existence of a family of processes ( $X^\kappa; \kappa > 0$ ) defined on the same probability space, satisfying pathwise the previous equality, and with the right marginal distributions (reflected, killed Lévy processes).

All results stated in the remainder of the paper hold in the supercritical case if we replace the Lévy process  $X$  by the projective limit of  $X^\kappa$  as  $\kappa \rightarrow \infty$ . More simply, we can construct the same consistent family ( $e^\kappa; \kappa > 0$ ) for the excursion of  $X - I$  away from 0. Then it is sufficient to replace each excursion of  $X - I$  drifting to  $\infty$  (there are finitely many of them on any compact interval of local time) by a copy of  $e^\kappa$  for some sufficiently large  $\kappa$ . More precisely, for the excursion corresponding to an infimum equal to  $-x$  ( $x$  is the index of the excursion in the local time scale), we need that the modified excursion  $e^\kappa$  be such that all heights below  $h := L^0(x)$  be visited. In other words, one has to choose  $\kappa$  large enough so that for any  $\kappa' > \kappa$ , the occupation measure of the height process of  $e^{\kappa'}$  restricted to  $[0, h]$  remains equal to that of  $e^\kappa$ .

In order not to overload the reader with technicalities that are away from the core question of this paper, we chose not to develop this point further, and we leave it to the reader to modify the proofs of the next subsection in the obvious way at points where the supercritical case has to be distinguished.

**4.3. Law of the coalescent point process.** Now that we have the height process for the infinite CSB tree with an arbitrarily old genealogy and the ancestral coalescence times with multiplicities describing genealogy of its standing population, we proceed to describe their law. Furthermore, as was our main goal, we define an analogue of the coalescent point process with multiplicities ( $B_i; i \geq 0$ ) for the CSB tree.

Our results show similarities with the results of Duquesne and Le Gall in [8], Section 2.7, for the law of the reduced tree under the measure  $N(\cdot \mid \sup H \geq T)$ . However, our presentation is quite different for a number of reasons. First, we do not characterize the branching times in terms of the Markov kernel of the underlying CSB process, but rather treat these times as a sequence. Second, the branching tree is allowed to be supercritical. Third, the tree may have an infinite past.

Recall the definition of coalescence times of individuals in the discretized standing population  $(A_i^\varepsilon; i \geq 1)$  and their multiplicities  $(N_i^\varepsilon; i \geq 1)$  from (6) and (7).

**THEOREM 4.1.** *The joint law of  $(A_1^\varepsilon, N_1^\varepsilon)$  is given by*

$$(8) \quad \mathbb{P}(A_1^\varepsilon > x) = \frac{\tilde{\psi}(v(x))}{\tilde{\psi}(v(\varepsilon))}, \quad x \geq \varepsilon,$$

where  $\tilde{\psi}(u) := \psi(u)/u$  for any  $u > \eta$ . In addition, for any  $n \geq 2$  and  $x \geq \varepsilon$ ,

$$\mathbb{P}(A_1^\varepsilon \in dx, N_1^\varepsilon = n)/dx = \mathbb{P}(A_1^\varepsilon > x)v(x)^n \int_{(0,\infty)} \Lambda(dz)e^{-v(x)z} \frac{z^{n+1}}{(n+1)!},$$

whereas for any  $x \geq \varepsilon$ ,

$$\mathbb{P}(A_1^\varepsilon \in dx, N_1^\varepsilon = 1)/dx = \mathbb{P}(A_1^\varepsilon > x)v(x) \left( \beta + \int_{(0,\infty)} \Lambda(dz)e^{-v(x)z} \frac{z^2}{2} \right).$$

**REMARK 3.** The formula giving the joint law of  $(A_1^\varepsilon, N_1^\varepsilon)$  can also be expressed (see the calculations in the proof) as follows:

$$\begin{aligned} &\mathbb{P}(A_1^\varepsilon \in dx, N_1^\varepsilon = n)/dx \\ &= \mathbb{P}(A_1^\varepsilon > x) \left( \beta \mathbf{1}_{\{n=1\}} + \int_{(0,+\infty)} \pi^{(x)}(dr)e^{-rv(x)} \frac{(rv(x))^n}{n!} \right), \end{aligned}$$

showing that when  $\beta = 0$ ,  $N_1^\varepsilon$  actually follows a mixed Poisson distribution.

**REMARK 4.** Similarly, as in the discrete case, observe that for subcritical trees [ $\eta = 0$  and  $\psi'(0+) > 0$ ], coalescence times can take the value  $+\infty$ , corresponding to the delimitation of quasi-stationary subpopulations with different (infinite) ancestor lineages. Indeed, from the previous statement,

$$\mathbb{P}(A_1^\varepsilon = +\infty) = \frac{\psi'(0+)}{\tilde{\psi}(v(\varepsilon))}.$$

In addition, the event  $\{A_1^\varepsilon = +\infty\}$  is the event that the whole (quasi-stationary) population has coalesced by time  $\varepsilon$ . In other words, if  $V$  denotes the coalescence time of a quasi-stationary population, also referred to as the time to most recent common ancestor, then

$$\mathbb{P}(V > x) = \frac{\psi'(0+)}{\tilde{\psi}(v(x))}.$$

In the discrete case, the study of  $V$  will be done in Section 5.2, in a slightly more detailed statement Proposition 5.2.

**PROOF OF THEOREM 4.1.** First notice that  $S_1$  is also the first hitting time of  $-Y_\varepsilon^0$  by  $X$ . Denote by  $(s, e_s)$  the excursion process of  $X - I$  away from 0, where  $-I$  serves as local time. Then the event  $\{A_1^\varepsilon > x\}$  is the event that the excursions  $(e_s; Y_\varepsilon^0 \leq s \leq Y_x^0)$  all have  $\sup H(e_s) - L_s^0 < 0$ . As a consequence,

$$\mathbb{P}(A_1^\varepsilon > x) = \mathbb{E} \exp - \sum_{Y_\varepsilon^0 \leq s \leq Y_x^0} \chi_{\{\sup H(e_s) < L_s^0\}},$$

where  $\chi_{\{A\}} = 0$  on  $A$  and  $\chi_{\{A\}} = +\infty$  on  ${}^cA$ , so that using the exponential formula for the excursion point process,

$$\mathbb{P}(A_1^\varepsilon > x) = \mathbb{E} \exp - \int_{[Y_\varepsilon^0, Y_x^0]} ds N(\sup H > L_s^0) = \mathbb{E} \exp - \int_{[Y_\varepsilon^0, Y_x^0]} ds v(L_s^0).$$

Now, since  $L^0$  is the right-inverse of  $Y^0$ , and because of the definition of  $Y^0$  (Definition 3.3) in terms of the Lebesgue measure and the Poisson point process  $(u, \Delta_u)$ , we get

$$\int_{[Y_\varepsilon^0, Y_x^0]} ds v(L_s^0) = \beta \int_{[\varepsilon, x]} du v(u) + \sum_{\varepsilon \leq u \leq x} v(u) \Delta_u,$$

so that

$$\begin{aligned} (9) \quad & \mathbb{P}(A_1^\varepsilon > x) \\ &= \exp - \left( \beta \int_{[\varepsilon, x]} du v(u) + \int_{[\varepsilon, x]} du \int_{[0, \infty)} \pi^{(u)}(dr) (1 - e^{-rv(u)}) \right). \end{aligned}$$

We compute the second term inside the exponential thanks to the Fubini–Tonelli theorem as

$$\begin{aligned} \int_{[0, \infty)} \pi^{(u)}(dr) (1 - e^{-rv(u)}) &= \int_{[0, \infty)} dr e^{v(u)r} \int_{(r, \infty)} e^{-v(u)z} \Lambda(dz) (1 - e^{-rv(u)}) \\ &= \int_{(0, \infty)} \Lambda(dz) e^{-v(u)z} \int_{(0, z)} dr (e^{v(u)r} - 1) \\ &= \frac{1}{v(u)} \int_{(0, \infty)} \Lambda(dz) e^{-v(u)z} (e^{v(u)z} - 1 - v(u)z). \end{aligned}$$

Finally, we get

$$\mathbb{P}(A_1^\varepsilon > x) = \exp - \int_{[\varepsilon, x]} du F \circ v(u),$$

where

$$F(\lambda) := \beta\lambda + \frac{1}{\lambda} \int_{(0, \infty)} \Lambda(dz) (1 - e^{-\lambda z} - \lambda z e^{-\lambda z}), \quad \lambda > \eta.$$

But elementary calculus shows that

$$F(\lambda) = \psi'(\lambda) - \tilde{\psi}(\lambda), \quad \lambda > \eta,$$

so that by the change of variable  $y = v(u)$ ,  $u = \phi(y)$ , we get

$$\begin{aligned} \mathbb{P}(A_1^\varepsilon > x) &= \exp - \int_{[v(x), v(\varepsilon)]} \frac{dy}{\psi(y)} F(y) \\ &= \exp - \int_{[v(x), v(\varepsilon)]} dy \left( \frac{\psi'(y)}{\psi(y)} - \frac{1}{y} \right) \\ &= \frac{\tilde{\psi}(v(x))}{\tilde{\psi}(v(\varepsilon))}, \end{aligned}$$

which shows the first part of the theorem.

As for the second part, the event  $\{A_1^\varepsilon \in dx, N_1^\varepsilon = n\}$  is the event that the excursions  $(e_s; Y_\varepsilon^0 \leq s \leq Y_{x-dx}^0)$  all have  $\sup H(e_s) - L_s^0 < 0$ , and  $n$  is the number of excursions  $(e_s; Y_{x-dx}^0 \leq s \leq Y_x^0)$  (for which  $L_s^0 \in dx$ ) such that  $\sup H(e_s) \geq x$ . Next, notice that  $Y_x^0 - Y_{x-dx}^0 = \beta dx + (Y_x^0 - Y_{x-}^0) \mathbf{1}_{\{E^0(dx)\}}$ , where  $E^0(dx)$  is the event that  $Y^0$  has a jump in the interval  $(x - dx, x)$ . As a consequence, by the compensation formula applied to the Poisson point process of jumps of  $Y^0$ ,

$$\mathbb{P}(A_1^\varepsilon \in dx, Y_x^0 - Y_{x-}^0 \in dr, N_1^\varepsilon = n) = \mathbb{P}(A_1^\varepsilon > x) dx \pi^{(x)}(dr) e^{-rv(x)} \frac{(rv(x))^n}{n!},$$

since  $N(\sup H \geq x) = v(x)$ . Also

$$\mathbb{P}(A_1^\varepsilon \in dx, Y_x^0 = Y_{x-}^0, N_1^\varepsilon = n) = 0,$$

if  $n \geq 2$ , whereas

$$\mathbb{P}(A_1^\varepsilon \in dx, Y_x^0 = Y_{x-}^0, N_1^\varepsilon = 1) = \mathbb{P}(A_1^\varepsilon > x) \beta dx v(x).$$

As a consequence, for any  $n \geq 2$ ,

$$\begin{aligned} \mathbb{P}(A_1^\varepsilon \in dx, N_1^\varepsilon = n) &= dx \mathbb{P}(A_1^\varepsilon > x) \frac{v(x)^n}{n!} \int_{[0, \infty)} \pi^{(x)}(dr) e^{-rv(x)} r^n \\ &= dx \mathbb{P}(A_1^\varepsilon > x) \frac{v(x)^n}{n!} \int_{[0, \infty)} dr r^n \int_{(r, \infty)} \Lambda(dz) e^{-zv(x)} \\ &= dx \mathbb{P}(A_1^\varepsilon > x) v(x)^n \int_{(0, \infty)} \Lambda(dz) e^{-v(x)z} \frac{z^{n+1}}{(n+1)!}, \end{aligned}$$

which is the desired result. The last result can be obtained by the same calculation. Summing over  $n$  yields

$$\begin{aligned} \mathbb{P}(A_1^\varepsilon \in dx) &= dx \mathbb{P}(A_1^\varepsilon > x) \\ &\quad \times \left( \beta v(x) + \frac{1}{v(x)} \int_{(0, \infty)} \Lambda(dz) (1 - e^{-v(x)z} - v(x)z e^{-v(x)z}) \right) \\ &= dx \mathbb{P}(A_1^\varepsilon > x) F \circ v(x) \end{aligned}$$



as expected, since  $\mathbb{P}(A_1^\varepsilon > x) = \exp - \int_{[\varepsilon, x]} du F \circ v(u)$ .  $\square$

Finally, we define the coalescent point process with multiplicities for the  $\varepsilon$ -discretized standing population. At the end of Section 2.2 we gave an alternative characterization of  $(B_k; k \geq 0)$  for the BGW tree as a sum of point masses [see (1)] whose multiplicities record the number of their future appearances as coalescent times, until the first future appearance of a larger coalescence time. Since in the CSB tree we do not have a process analogous to  $(D_i; i \geq 1)$ , we actually define the  $\varepsilon$ -discretized coalescent point process with multiplicities from this angle.

For each  $i \geq 1$  and  $k \geq i$  we define  $N_{ik}^\varepsilon$  as the residual multiplicity of  $A_i^\varepsilon$  at time  $k$ ,

$$N_{ik}^\varepsilon := \#\{j \geq k : A_j^\varepsilon = A_i^\varepsilon\},$$

so  $N_{ii}^\varepsilon = N_i^\varepsilon$ . Next define for each  $k \geq 1$  the random finite point measure  $B_k^\varepsilon$  on  $(\varepsilon, +\infty]$  as

$$B_k^\varepsilon := \sum_{i \leq k} N_{ik}^\varepsilon \delta_{A_i^\varepsilon} \mathbf{1}_{\{A_i^\varepsilon < A_{i'}^\varepsilon, \forall i' < i : N_{i'k}^\varepsilon \neq 0\}},$$

where  $\delta$  denotes a Dirac measure, and by convention let  $B_0^\varepsilon$  be the null measure.

Recall that  $\mathfrak{s}$  is a mapping from a point measure on  $\mathbb{R}$  to the minimum of its support. The following result provides the law for the coalescent point process with multiplicities of the  $\varepsilon$ -discretized population in the CSB tree.

**THEOREM 4.2.** *The sequence of finite point measures  $(B_i^\varepsilon; i \geq 0)$  is a Markov chain started at the null measure. For any finite point measure  $b = \sum_{j \geq 1} n_j \delta_{a_j}$ , with  $n_j \in \mathbb{N}$  and  $a_j \in \mathbb{R}^+$ , the law of  $B_{k+1}^\varepsilon$  conditionally given  $B_k^\varepsilon = b$ , is given by the following transition. Let  $a_1 := \mathfrak{s}(b)$ ,  $b^* := b - \delta_{a_1}$ ,  $a_1^* := \mathfrak{s}(b^*)$ . Let  $(A, N)$  be a r.v. with values in  $(\varepsilon, +\infty] \times \mathbb{N}$  distributed as  $(A_1^\varepsilon, N_1^\varepsilon)$ . Then*

$$B_{k+1}^\varepsilon := \begin{cases} b^* + N \delta_A, & \text{if } A < a_1^*, \\ b^*, & \text{otherwise.} \end{cases}$$

*In addition, as in the discrete case, the coalescent point process can be deduced from  $B^\varepsilon$  as*

$$A_k^\varepsilon = \mathfrak{s}(B_k^\varepsilon).$$

**REMARK 5.** When  $X$  is a diffusion, the measure  $\rho^0$  has no atoms, so there are no repeats of coalescence times ( $N = 1$  a.s.). In this case, at each step of the chain there is only one nonzero mass, that is,  $\forall k \geq 1, B_k^\varepsilon = \delta_{A_k^\varepsilon}$ . The previous statement shows that the sequence starts with  $B_1^\varepsilon = \delta_{A_1^\varepsilon}$  where  $A_1^\varepsilon$  is distributed as  $A$ , and in every transition the single point mass from the previous step is erased and a new point mass  $\delta_{A_{k+1}^\varepsilon}$  is added with independently chosen  $A_{k+1}^\varepsilon$  distributed as  $A$ . So the random variables  $(A_i^\varepsilon)_{i \geq 1}$  are i.i.d.

On the other hand, when  $X$  is a diffusion (and only in that case; see [8]), the height process  $H$  is a Markov process. The coalescence times  $(A_i^\varepsilon; i \geq 1)$  are just the depths of excursions of  $H$  (with depth greater than  $\varepsilon$ ) below some fixed level. This again implies that the  $(A_i^\varepsilon)_{i \geq 1}$  are i.i.d. Moreover, by taking  $\varepsilon \rightarrow 0$  in equation (8), one can compute the intensity measure  $\mu$  of the Poisson point process of excursion depths [its tail is given by  $\tilde{\mu}(x) = \tilde{\psi}(v(x))$ ]. In particular, in the Brownian case, it is known that the height process is (reflected) Brownian, so that  $\mu(dx)$  is proportional to  $x^{-2} dx$  (see [1, 26]).

For the BGW tree, the discrete analogue of the height process  $H$  is, again, in general not a Markov process, except in special cases of the offspring distribution, namely, the only exceptions are when  $\xi$  is linear fractional (see Section 5.1 for definition). In these cases we will also observe that the coalescence times  $(A_i; i \geq 1)$  are i.i.d. (see Proposition 5.1).

REMARK 6. Even in general, that is, in the presence of multiplicities, we know that there exists a coalescent point process whose truncation at level  $\varepsilon$  is the sequence  $(A_i^\varepsilon; i \geq 1)$ . It is the process of excursion depths of the height process in the local time scale. However, the question of characterizing the distribution of this coalescent point process (without truncation) remains open. A natural idea would be to use a Poisson point process with intensity measure  $dt \sum_{n \geq 1} v(dx, n) \delta_n$ , where

$$v(dx, \{n\})/dx = \tilde{\psi}(v(x))v(x)^n \left( \beta \mathbf{1}_{\{n=1\}} + \int_{(0, \infty)} \Lambda(dz) e^{-v(x)z} \frac{z^{n+1}}{(n+1)!} \right).$$

Indeed, since  $(A_1^\varepsilon, N_1^\varepsilon)$  has the same law as the first atom of this point process conditional on its first component being larger than  $\varepsilon$ , it is easy to embed the sequence  $(A_i^\varepsilon, i \geq 1)$  into the atoms of this Poisson point process by truncating atoms larger than those whose multiplicity has not yet been exhausted. However, preliminary calculations indicate that convergence of this new point process as  $\varepsilon \rightarrow 0$  is not evident.

PROOF OF THEOREM 4.2. We need to introduce some notation. For any stochastic process  $W$  admitting a height process in the sense of (4), we denote by  $H^W$  its height process, that is,

$$H_t^W := \lim_{k \rightarrow \infty} \frac{1}{\varepsilon_k} \int_0^t \mathbf{1}_{\{W_s < I_{W,t}^s + \varepsilon_k\}} ds,$$

where

$$I_{W,t}^s := \inf_{s \leq r \leq t} W_r.$$

Further, if  $W$  has finite lifetime, denoted  $T$ , we denote by  $\rho^W$  the great-aunt measure associated with  $W$ , in the sense that

$$\langle \rho^W, f \rangle := \int_{[0, T]} d_s I_{W,T}^s f(H_T^W - H_s^W).$$

In particular, if  $W$  is the Lévy process killed at  $\sigma_h$  under  $N(\cdot \mid \sup H > h)$ , then we know from Theorem 3.4 that  $\rho^W$  and the trace of  $\rho^0$  on  $[0, h]$  are equally distributed. Now if  $\mu$  is a positive measure on  $\mathbb{R}_+$  and  $L^\mu$  denotes the right-inverse of the nondecreasing function  $x \mapsto \mu([0, x])$ , and if  $I^W$  denotes the past infimum of the shifted path  $W - W_0$ , we denote by  $\Phi(\mu, W)$  the generalized height process

$$\Phi_t(\mu, W) := H_t^W - L^\mu(-I_t^W).$$

In particular, in the (sub)critical case,  $H^*$  is distributed as  $\Phi(\rho^0, X)$ . Finally, recall that  $k$  denotes the killing operator and  $\theta$  the shift operator, in the sense that  $X \circ \theta_t = (X_{t+s}; s \geq 0)$  and  $X \circ k_t$  is  $(X_{s \wedge t}; s \geq 0)$ . For any path  $(X_s; s \geq 0)$  and any positive real number  $t$ , if  $W := X \circ k_t$  and  $W' = X \circ \theta_t$ , then it is not hard to see that for any  $s \geq 0$ ,

$$(10) \quad H_{t+s}^X = \Phi_s(\rho^W, W') + H_t^W.$$

In particular, applying this to  $X$  under  $N(\cdot \mid \sup H > h)$  and to  $t = \sigma_h$ , the strong Markov property at  $\sigma_h$  yields

$$(11) \quad H^X \circ \theta_{\sigma_h} \stackrel{d}{=} \Phi(\rho, X \circ \theta_{\sigma_h}) + h,$$

where  $\rho$  is a copy of  $\rho^0$  independent of  $X$ .

Now we work conditionally on  $(A_1^\varepsilon, N_1^\varepsilon) = (a, n)$ . Recall that  $S_1 = \inf\{t \geq 0: H_t^* = -\varepsilon\}$  and  $T_1 = \inf\{t \geq S_1: H_t^* = 0\}$ . Set  $V_1 := T_1$  and define recursively for  $i = 1, \dots, n$ ,

$$\begin{aligned} U_i &:= \sup\{t < V_i: H_t^* = -a\}, \\ W_i &:= \inf\{t > V_i: H_t^* = -a\}, \\ V_{i+1} &:= \inf\{t > W_i: H_t^* = 0\}. \end{aligned}$$

Last, define

$$W_0 := \inf\{t > 0: X_t = -Y_{a-}^0\} \quad \text{and} \quad U_{n+1} := \inf\{t > 0: X_t = -Y_a^0\}.$$

As seen in the proof of Theorem 4.1, the subpaths  $e^i := (X_{t+U_i} - X_{U_i}; 0 \leq t \leq W_i - U_i)$  are the  $n$  excursions of  $X$  above its infimum whose height reaches level  $a$ , so that for all  $t \in [W_0, U_{n+1}]$ ,  $I_t \in (-Y_{a-}^0, Y_a^0]$  and  $H_t^* = H_t - a$ . An application of the strong Markov property yields the independence of the  $n$  subpaths  $e^i$  and the fact that they are all distributed as  $N(\cdot \mid \sup H \geq a)$ . Also,  $X' := (X_{t+U_{n+1}} - X_{U_{n+1}}; t \geq 0)$  is a copy of  $X$  independent from all the previous subpaths and from  $Y^0$ . Notice that for all  $t \in [W_0, U_{n+1}]$ ,  $H_t^* \geq -a$ . As a consequence, by continuity of the height process,  $H^*$  takes the value  $-a$  at all points of the form  $U_i$  and  $W_i$ , it hits 0 only in intervals of the form  $[V_i, W_i)$  and takes values in  $[-a, 0)$  on all intervals of the form  $[W_i, U_{i+1}]$ . As a consequence, if we excise all the paths of  $H^*$  on intervals of the form  $(W_i, U_{i+1})$ ,  $i = 0, \dots, n$ , we will still get the same coalescent point process. Also notice that those paths

are independent of the  $n$  excursions  $e^i$  and only depend on  $Y^0$  through  $\Delta Y_a^0$ . As a consequence, recalling the notation in Section 4.2, if  $\mathcal{A}^\varepsilon$  denotes the mapping that takes a height process to its sequence of  $\varepsilon$ -discretized coalescent levels, that is,  $\mathcal{A}^\varepsilon : H^\star \mapsto (A_1^\varepsilon, A_2^\varepsilon, \dots)$ , where  $A_1^\varepsilon, A_2^\varepsilon, \dots$  are defined by (6), then we have

$$\mathcal{A}^\varepsilon(H^\star) = (a, \mathcal{A}^\varepsilon(H(e^1) - a), a, \dots, \mathcal{A}^\varepsilon(H(e^{n-1}) - a), a, \mathcal{A}^\varepsilon(H_n'')),$$

where  $H_n''$  is the concatenation of  $H(e^n) - a$  and of  $\Phi(\rho^0 \circ \theta_a, X')$ , writing  $\rho^0 \circ \theta_a$  for the measure associated with the jump process  $(Y_{s+a}^0 - Y_a^0; s \geq 0)$ . First observe that, as long as the coalescent point process is concerned, we can again excise the parts of each of the  $n$  subpaths in the previous display going from height  $-a$  to height 0. This amounts to considering excursions  $e^i$  only from the first time  $\sigma_a$  they reach height  $a$ . But recall from (11) that  $H(e) \circ \theta_{\sigma_a} - a \stackrel{d}{=} \Phi(\rho, e \circ \theta_{\sigma_a})$ , which is distributed as the process  $H^\star$  killed upon reaching  $-a$ . Second, by the same argument as stated previously, erasing the part of  $H_n''$  before its first hitting time of 0, we get the concatenation of a copy of  $H^\star$  killed upon reaching  $-a$  and of  $\Phi(\rho^0 \circ \theta_a, X')$ , where we remember that  $X'$  is an independent copy of  $X$ . The result has the law of  $\Phi(\rho^0, X)$ , that is, it has the law of  $H^\star$ . In conclusion, this gives the following conditional equality in distribution:

$$\mathcal{A}^\varepsilon(H^\star) = (a, \mathcal{A}^\varepsilon(H_1^\star), a, \dots, \mathcal{A}^\varepsilon(H_{n-1}^\star), a, \mathcal{A}^\varepsilon(H_n'')),$$

where  $H_n''$  is a copy of  $H^\star$  and the  $H_i^\star$  are independent copies of  $H^\star$  killed upon reaching  $-a$ , all independent of  $H_n''$ .

Now observe that since the law of  $A_1^\varepsilon$  is absolutely continuous, the branch length  $a$  will occur exactly  $n$  times [in particular, it will not appear in  $\mathcal{A}^\varepsilon(H_n'')$ ]. Also, because all the heights between successive occurrences of  $a$  are smaller than  $a$ , the following conditional equality can be inferred from the previous display and the definition of the sequence  $B^\varepsilon := (B_i^\varepsilon)$  of point measures

$$B^\varepsilon = (0, n\delta_a, (n-1)\delta_a + B^1, (n-2)\delta_a, \dots, \delta_a + B^{n-1}, \delta_a, B'),$$

where  $B'$  is a copy of  $B^\varepsilon$  and the  $B^i$  are independent copies of the sequence of point measures associated with a coalescent point process  $A^\varepsilon$  killed at its first value greater than  $a$  (in the usual sense that this value is not included in the killed sequence). In passing, an induction argument on the cardinal of the support of  $B^\varepsilon$  shows that  $A^\varepsilon = \mathfrak{s}(B^\varepsilon)$ . Also, each  $B^i$  is reduced to a sequence of length 1 with single value 0, with probability  $\mathbb{P}(A^\varepsilon > a)$ . Otherwise, it starts in the state  $N\delta_a$ , where  $(A, N)$  has the law of  $(A_1^\varepsilon, N_1^\varepsilon)$  conditional on  $A_1^\varepsilon < a$ . Note that the sequence following the state  $\delta_a$  [which corresponds to the case when  $\mathfrak{s}(b) \neq \mathfrak{s}(b^*)$ ], has the same law as  $B^\varepsilon$ , dropping the dependence upon  $a$ .

Now assume that  $B_k^\varepsilon = \sum_{j \geq 1} n_j \delta_{a_j}$ . The following conclusions follow by induction from the last assertion. If  $n_1 \geq 2$  (i.e.,  $a_1^* = a_1$ ) and  $K$  is the first time after  $k+1$  that  $a_1$  has its multiplicity decreased, then the sequence  $(B_i^\varepsilon; k+1 \leq i < K)$  is independent of  $(B_i^\varepsilon; i \leq k)$  and has the law of the sequence associated with a

coalescent point process killed at its first value greater than  $a_1$ . If  $n_1 = 1$  (i.e.,  $a_1^* = a_2$ ) and  $K$  is the first time (after  $k$ ) that  $a_2$  has its multiplicity decreased, then the sequence  $(B_i^\varepsilon; k + 1 \leq i < K)$  is independent of  $(B_i^\varepsilon; i \leq k)$  and has the law of the sequence associated with a coalescent point process killed at its first value greater than  $a_2$ . In particular,  $B_{k+1}^\varepsilon = B_k^\varepsilon - \delta_{a_1}$  with probability  $\mathbb{P}(A^\varepsilon > a_1^*)$ , and  $B_{k+1}^\varepsilon = B_k^\varepsilon - \delta_{a_1} + N\delta_A$ , with probability  $\mathbb{P}(A^\varepsilon < a_1^*)$ , where  $(A, N)$  has the law of  $(A_1^\varepsilon, N_1^\varepsilon)$  conditional on  $A_1^\varepsilon < a_1^*$ .  $\square$

4.4. *Convergence of the coalescent point process.* We now present the theorem that connects the discrete case coalescent process, based on the offspring distribution  $\xi_p$ , to the continuous case coalescent process, based on the associated branching mechanism  $\psi$ . Let us assume, as in Section 3.4, that for some sequence  $(\gamma_p, p \geq 1)$ ,  $\gamma_p \rightarrow \infty$  as  $p \rightarrow \infty$  and a sequence of r.v.  $(\xi_p, p \geq 1)$  such that the rescaled BGW process started at  $[px]$  with offspring distribution  $\xi_p$  when re-scaled to  $p^{-1}Z_{[\gamma_p \cdot]}^{(p)}$ , converges in Skorokhod space to a CSB process  $Z$  with branching mechanism  $\psi$  started at  $x$ .

In order to obtain convergence for the coalescent processes we need to define a discretized version of the discrete case process by considering only the individuals whose coalescent times are greater than  $\gamma_p\varepsilon$  for some fixed  $\varepsilon > 0$ . Recall that for a point measure  $b$ ,  $\mathfrak{s}(b)$  denotes the minimum of its support. Start with the sequence of finite point measures  $(B_i; i \geq 0)$  whose law is given by Theorem 2.2. Let  $\tau_0 := 0$ , and for any  $i \geq 1$ , let

$$\tau_i := \inf\{n \geq \tau_{i-1} : \mathfrak{s}(B_n) \geq \gamma_p\varepsilon\}.$$

Define the  $\gamma_p\varepsilon$ -discretized process of point measures

$$(B_i^{p,\varepsilon}; i \geq 0) := (B_{\tau_i}; i \geq 0).$$

Let  $\mathcal{B} = \{\sum_{i=1}^n b_i \delta_{a_i} : n \in \mathbb{N}, b_i \in \mathbb{N}, a_i \in \mathbb{R}_+\}$  be the space of all finite point mass measures on  $\mathbb{R}_+$ , equipped with the usual vague topology. Let  $\mathcal{R}^p : \mathcal{B} \mapsto \mathcal{B}$  be a function re-scaling the point mass measures so that

$$\mathcal{R}^p : \sum_{i=1}^n b_i \delta_{a_i} \mapsto \sum_{i=1}^n b_i \delta_{\gamma_p^{-1} a_i}.$$

Since  $(B_i^{p,\varepsilon}; i \geq 0)$  is a Markov chain on  $\mathcal{B}$ , it is a random element of  $\mathcal{B}^{\{0,1,2,\dots\}}$ , equipped with the product of vague topologies on  $\mathcal{B}$ .

**THEOREM 4.3.** *The sequence of re-scaled discretized Markov chains  $(\mathcal{R}^p(B_i^{p,\varepsilon}); i \geq 0)$  converges in distribution on the space  $\mathcal{B}^{\{0,1,2,\dots\}}$  to the Markov chain  $(B_i^\varepsilon; i \geq 0)$  whose law is given by Theorem 4.2 as  $p \rightarrow \infty$ .*

PROOF. Note that the initial values for the sequence  $\mathcal{R}^p(B_0^{p,\varepsilon})$ , as well as for the limit  $B_0^\varepsilon$ , are simply null measures. In order to describe the transition law of the discretized process  $(B_i^{p,\varepsilon}; i \geq 0)$ , condition on its value at step  $i$

$$B_{\tau_i} = b \quad \text{with } a_1 = \mathfrak{s}(b) \geq \gamma_p \varepsilon.$$

Condition further on the values of  $\tau_i$  and  $\tau_{i+1}$  for the unsampled process  $(B_k; k \geq 1)$ . With  $b^* = b - \delta_{a_1}$ , we have that for all  $1 \leq j \leq \tau_{i+1} - \tau_i - 1$ ,

$$B_{\tau_i+j} = b^* + \sum_{j \geq 1} n_j \delta_{a_j} \quad \text{with } a_j < \gamma_p \varepsilon, \forall j \geq 1$$

since, by definition of  $\tau_{i+1}$  for  $1 \leq j \leq \tau_{i+1} - \tau_i - 1$ , the smallest mass in  $B_{\tau_i+j}$  must be smaller than  $\gamma_p \varepsilon$ , and in each step only the weight of the smallest mass is decreased.

On the event  $\tau_{i+1} = \tau_i + 1$ , the transition rule from Theorem 2.2 for step  $\tau_i$  to  $\tau_{i+1}$  gives that, for  $a_1^* = \mathfrak{s}(b^*)$ ,

$$(12) \quad B_{\tau_{i+1}} = \begin{cases} b^* + N^{p,\varepsilon} \delta_{A^{p,\varepsilon}}, & \text{if } A^{p,\varepsilon} < a_1^* \text{ and } A^{p,\varepsilon} \neq a_1, \\ b^*, & \text{otherwise,} \end{cases}$$

where  $(A^{p,\varepsilon}, N^{p,\varepsilon})$  is distributed as  $(A_1, \zeta'_{A_1})$  conditional on  $A_1 \geq \gamma_p \varepsilon$ . The conditioning in this law follows from the definition of  $\tau_{i+1}$  as the first time after  $\tau_i$  for which  $\mathfrak{s}(B_{i+1}^{p,\varepsilon}) \geq \gamma_p \varepsilon$ .

On the event  $\tau_{i+1} > \tau_i + 1$ , since  $\mathfrak{s}(B_{i+1}^{p,\varepsilon}) \geq \gamma_p \varepsilon$ , by step  $\tau_{i+1} - 1$  all of the masses from  $\sum_{j \geq 1} n_j \delta_{a_j}$  must have been eliminated except for one mass  $a$  that is smaller than  $\gamma_p \varepsilon$  whose weight at this step is 1, so

$$B_{\tau_{i+1}-1} = b^* + \delta_a \quad \text{with } a < \gamma_p \varepsilon.$$

Since the smallest mass in  $B_{\tau_{i+1}-1}$  is  $\delta_a$ , the transition rule from Theorem 2.2 for step  $\tau_{i+1} - 1$  to  $\tau_{i+1}$  gives that  $B_{\tau_{i+1}-1}^* = B_{\tau_{i+1}-1} - \delta_a = b^*$ , so  $\mathfrak{s}(B_{\tau_{i+1}-1}^*) = \mathfrak{s}(b^*) = a_1^*$ . Also, a new mass  $(A, N)$  is added only if  $A < a_1^*$  and  $A \neq a = \mathfrak{s}(b^* + \delta_a)$ . Note that we must also have  $A \geq \gamma_p \varepsilon$  as in the next step  $\mathfrak{s}(B_{\tau_{i+1}}) \geq \gamma_p \varepsilon$ , so the added mass is again distributed as  $(A_1, \xi'_{A_1})$  conditional on  $A_1 \geq \gamma_p \varepsilon$ . Integrating over possible values for  $B_{\tau_{i+1}-1}$ ,  $\tau_i$ , and  $\tau_{i+1}$ , we have that, conditionally given  $B_i^{p,\varepsilon} = b$ , the transition rule for  $B_{i+1}^{p,\varepsilon}$  is

$$B_{\tau_{i+1}} = \begin{cases} B_{\tau_i}^* + N^{p,\varepsilon} \delta_{A^{p,\varepsilon}}, & \text{if } A^{p,\varepsilon} < a_1^* \text{ and } A^{p,\varepsilon} \neq a_1, \\ B_{\tau_i}^*, & \text{otherwise,} \end{cases}$$

where  $(A^{p,\varepsilon}, N^{p,\varepsilon})$  is distributed as  $(A_1, \zeta'_{A_1})$  conditional on  $A_1 \geq \gamma_p \varepsilon$ .

For the re-scaled process  $\mathcal{R}^p(B^{p,\varepsilon})$ , the transition rule for  $\mathcal{R}^p(B_{i+1}^{p,\varepsilon})$ , conditional on the value of  $\mathcal{R}^p(B_i^{p,\varepsilon}) = \mathcal{R}^p(b)$ , is then

$$\mathcal{R}^p(B_{i+1}^{p,\varepsilon}) = \begin{cases} \mathcal{R}^p(b)^* + N^{p,\varepsilon} \delta_{\gamma_p^{-1} A^{p,\varepsilon}}, & \text{if } A^{p,\varepsilon} < a_1^* \text{ and } A^{p,\varepsilon} \neq a_1, \\ \mathcal{R}^p(b)^*, & \text{otherwise,} \end{cases}$$

where we define  $\mathcal{R}^p(a_1) := \mathfrak{s}(\mathcal{R}^p(b))$ ,  $\mathcal{R}^p(b)^* := \mathcal{R}^p(b) - \delta_{\mathcal{R}^p(a_1)}$  and  $\mathcal{R}^p(a_1^*) := \mathfrak{s}(\mathcal{R}^p(b)^*)$ .

If we can now show convergence in distribution of

$$(\gamma_p^{-1} A^{p,\varepsilon}, N^{p,\varepsilon}) \stackrel{d}{=} (\gamma_p^{-1} A_1, \zeta'_{A_1}) \mid A_1 \geq \gamma_p \varepsilon \xrightarrow{p \rightarrow \infty} (A_1^\varepsilon, N_1^\varepsilon),$$

then our claim on point process convergence will follow from the description of the transition rule for  $(B_i^\varepsilon; i \geq 1)$  from Theorem 4.2 and the standard convergence arguments for a sequence of Markov chains based on weak convergence of their initial values and transition laws.

We now express  $(A_1, \zeta'_{A_1})$  in terms of the great-aunt measure  $\rho^{(p)}$  of a single quasi-stationary BGW tree with offspring distribution  $\xi_p$ . Consider the  $(0, 1)$  individual in our doubly infinite embedding of quasi-stationary BGW genealogies. Recall (see Remark 2) that  $(\zeta'_n; n \geq 1)$  is a sequence of independent r.v.s which conditionally on  $(\rho_n^{(p)}, n \geq 1)$  are binomial with parameters  $\rho_n^{(p)}$  and  $p_{n-1} = \mathbb{P}(Z_n^{(p)} \neq 0 \mid Z_0^{(p)} = 1)$ .

First for  $A_1$ , take any  $x > \varepsilon$ , then

$$\begin{aligned} \mathbb{P}(A_1 > \gamma_p x \mid A_1 \geq \gamma_p \varepsilon) &= \prod_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p x \rfloor} \mathbb{P}(\zeta'_i = 0) = \mathbb{E} \prod_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p x \rfloor} (1 - p_{i-1})^{\rho_i^{(p)}} \\ &= \mathbb{E} \exp\left( \sum_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p x \rfloor} \rho_i^{(p)} \ln(1 - p_{i-1}) \right). \end{aligned}$$

Let  $\tau_{\text{ext}}^{(p)}$  denote the extinction time of a BGW process  $Z^{(p)}$  started with  $Z_0^{(p)} = p$ , then

$$1 - p_{n-1} = \mathbb{P}(Z_{n-1}^{(p)} = 0 \mid Z_0^{(p)} = 1) = \mathbb{P}(\tau_{\text{ext}}^{(p)} \leq n - 1)^{1/p}.$$

Let  $\tau_{\text{ext}}$  denote the extinction time of the CSB process  $Z$  started with  $Z_0 = 1$ , then the assumption  $\liminf_{p \rightarrow \infty} \mathbb{P}(Z_{\lfloor \delta \gamma_p \rfloor}^{(p)} = 0) > 0$  guarantees that whenever  $\gamma_p^{-1} i_p \rightarrow i$ , we have

$$(1 - p_{i_p-1})^p = \mathbb{P}(\gamma_p^{-1} \tau_{\text{ext}}^{(p)} \leq \gamma_p^{-1} i_p - \gamma_p^{-1}) \xrightarrow{p \rightarrow \infty} \mathbb{P}(\tau_{\text{ext}} < i) = e^{-v(i)}.$$

Let us define  $f^{\varepsilon,x}(\cdot) = v(\cdot) \mathbf{1}_{[\varepsilon,x]}(\cdot)$  and a sequence of functions  $\{f_p\}$  such that

$$f_{p,\gamma_p}^{\varepsilon,x}(\cdot) = f_p^{\varepsilon,x}(\gamma_p^{-1} \cdot) = -\ln(1 - p_{\cdot-1})^p \chi_p^{\varepsilon,x}(\cdot),$$

where  $\chi_p^{\varepsilon,x}$  is a sequence of bounded continuous functions approximating  $\mathbf{1}_{[\lceil \gamma_p \varepsilon \rceil, \lfloor \gamma_p x \rfloor]}$  converging pointwise to  $\mathbf{1}_{[\varepsilon,x]}(\cdot)$ . Then we have that  $f_p^{\varepsilon,x}(\gamma_p^{-1} i_p) \rightarrow$

$f^{\varepsilon,x}(i) = v(i)\mathbf{1}_{\varepsilon \leq i \leq x}$  as  $p \rightarrow \infty$  whenever  $\gamma_p^{-1}i_p \rightarrow i$ . By Theorem 3.4(iii) it follows that

$$\sum_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p x \rfloor} \frac{\rho_i^{(p)}}{p} \ln(1 - p_{i-1})^p \xrightarrow[p \rightarrow \infty]{d} -\langle \rho^0, v\mathbf{1}_{[\varepsilon,x]} \rangle$$

showing that

$$\mathbb{P}(A_1 > \gamma_p x \mid A_1 \geq \gamma_p \varepsilon) \xrightarrow[p \rightarrow \infty]{} \mathbb{E} \exp(-\langle \rho^0, v\mathbf{1}_{[\varepsilon,x]} \rangle).$$

By Definition 3.3 and equation (9) we have

$$\begin{aligned} & \mathbb{E} \exp(-\langle \rho^0, v\mathbf{1}_{[\varepsilon,x]} \rangle) \\ &= \exp\left(-\left(\beta \int_{[\varepsilon,x]} du v(u) + \int_{[\varepsilon,x]} du \int_{[0,\infty)} \pi^{(u)}(dr)(1 - e^{-rv(u)})\right)\right), \end{aligned}$$

which together with the results of Theorem 4.1 proves that  $\gamma_p^{-1}A_1 \mid A_1 > \gamma_p \varepsilon \xrightarrow{d} A_1^\varepsilon$ .

We next show that for any  $y > \varepsilon$  the sequence  $\sum_{i=\gamma_p \varepsilon}^{\gamma_p y} \zeta'_i$  converges as  $p \rightarrow \infty$  to a Cox point process on  $[\varepsilon, y]$  whose intensity measure given  $\rho^0$  is  $\langle \rho^0, v\mathbf{1}_{[\varepsilon,y]} \rangle$ . For any  $\lambda > 0$ ,

$$\begin{aligned} \mathbb{E} \exp\left(-\lambda \sum_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p y \rfloor} \zeta'_i\right) &= \mathbb{E} \prod_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p y \rfloor} (1 - p_{i-1} + p_{i-1}e^{-\lambda})^{\rho_i^{(p)}} \\ &= \mathbb{E} \exp\left(\sum_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p y \rfloor} \rho_i^{(p)} \ln(1 - p_{i-1} + p_{i-1}e^{-\lambda})\right). \end{aligned}$$

Let us define  $g^{\varepsilon,y}(\cdot) = v(\cdot)(1 - e^{-\lambda})\mathbf{1}_{[\varepsilon,y]}(\cdot)$  and a sequence of functions  $\{g_p^{\varepsilon,y}\}$  such that

$$g_{p,\gamma_p}^{\varepsilon,y}(\cdot) = g_p^{\varepsilon,y}(\gamma_p^{-1}\cdot) = -\ln(1 - p_{i-1} + p_{i-1}e^{-\lambda})^p \chi_p^{\varepsilon,y}(\cdot).$$

Since whenever  $\gamma_p^{-1}i_p \rightarrow i$ , we have

$$\begin{aligned} p_{i_p} &= \mathbb{P}(Z_{i_p}^{(p)} \neq 0 \mid Z_0^{(p)} = 1) = 1 - \mathbb{P}(\tau_{\text{ext}}^{(p)} \leq i_p)^{1/p} \\ &\approx 1 - \mathbb{P}(\tau_{\text{ext}} < \gamma_p^{-1}i_p)^{1/p} \approx v(\gamma_p^{-1}i_p)/p, \end{aligned}$$

it follows that  $g_p^{\varepsilon,y}(\gamma_p^{-1}i_p) \rightarrow g^{\varepsilon,y}(i) = v(i)(1 - e^{-\lambda})\mathbf{1}_{\varepsilon \leq i \leq y}$  as  $p \rightarrow \infty$  whenever  $\gamma_p^{-1}i_p \rightarrow i$ . By Theorem 3.4(iii),

$$\mathbb{E} \exp\left(-\lambda p^{-1} \sum_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p y \rfloor} \zeta'_i\right) \xrightarrow[p \rightarrow \infty]{} \mathbb{E} \exp(\langle \rho^0, v\mathbf{1}_{[\varepsilon,y]} \rangle (e^{-\lambda} - 1)),$$



showing that (e.g., [16], Theorem 16.29)

$$(13) \quad \sum_{i=\lceil \gamma_p \varepsilon \rceil}^{\lfloor \gamma_p y \rfloor} \zeta'_i \xrightarrow[p \rightarrow \infty]{d} \Xi_{[\varepsilon, y]},$$

where  $\Xi$  is a Cox point process whose intensity measure conditionally on  $\rho^0$  is  $\langle \rho^0, \nu \rangle$ . Finally, take  $\varepsilon < x' < x < x'' < y$ , such that  $\gamma_p(x' - x) \rightarrow 0$  and  $\gamma_p(x'' - x) \rightarrow 0$  as  $p \rightarrow \infty$ .

$$\begin{aligned} & \mathbb{P}(\gamma_p x' \leq A_1 \leq \gamma_p x'', \zeta'_{A_1} = n \mid A_1 > \gamma_p \varepsilon) \\ &= \mathbb{P}(\gamma_p x' \leq A_1 \leq \gamma_p x'' \mid A_1 > \gamma_p \varepsilon) \mathbb{P}(\zeta'_{A_1} = n \mid \gamma_p x' \leq A_1 \leq \gamma_p x'') \\ &\xrightarrow[p \rightarrow \infty]{} \mathbb{P}(A_1^\varepsilon \in dx) \mathbb{P}(\Xi_{[\varepsilon, y]}(dx) = n \mid \Xi_{[\varepsilon, y]}(dx) \neq \emptyset). \end{aligned}$$

Evaluating the probability that the point process with intensity measure  $\langle \rho^0, \nu \rangle$  takes on values  $n = 1$  and  $n \geq 2$  at height  $x$ , gives precisely the formulae given in Theorem 4.1, showing that for all  $n \geq 1$  and  $x > \varepsilon$ ,

$$\mathbb{P}(A_1 \in \gamma_p dx, \zeta'_{A_1} = n \mid A_1 > \gamma_p \varepsilon) \xrightarrow[p \rightarrow \infty]{} \mathbb{P}(A_1^\varepsilon \in dx, N_1^\varepsilon = n),$$

and the proof is complete.  $\square$

**5. Two applications in the discrete case.** In this section, we come back to the discrete case to display two further results on the coalescent point process.

5.1. *The linear-fractional case.* A BGW process is called *linear fractional* if there are two probabilities  $a$  and  $b$  such that

$$f(s) = a + \frac{(1 - a)(1 - b)s}{1 - bs}, \quad s \in [0, 1].$$

In other words,  $\xi$  is product of a Bernoulli r.v. with parameter  $(1 - a)$  and a geometric r.v. with parameter  $(1 - b)$  conditional on being nonzero. The expectation  $m$  of  $\xi$  is equal to

$$m = \frac{1 - a}{1 - b},$$

so that this BGW process is (sub)critical iff  $a \stackrel{(>)}{=} b$ . In general, the coalescent point process  $(A_i, i \geq 1)$  is not itself a Markov process, but in the linear-fractional case, it is a sequence of i.i.d. random variables. An alternative formulation of this observation was previously derived in [27].

**PROPOSITION 5.1.** *In the linear-fractional case with parameters  $(a, b)$ , the branch lengths of the coalescent  $(A_i, i \geq 1)$  are i.i.d. with distribution given by*

$$\mathbb{P}(A_1 > n) = \frac{b - a}{bm^n - a},$$

when  $a \neq b$ , and when  $a = b$  (critical case), by

$$\mathbb{P}(A_1 > n) = \frac{1 - a}{na + 1 - a}.$$

PROOF. Recall from Theorem 2.1 that  $A_i = \min\{n \geq 1 : D_i(n) \neq 0\}$ , so in particular, we can set  $A_0 := +\infty$ . We prove by induction on  $i \geq 1$  the following statement  $(\mathcal{S}_i)$ . The random variables  $(D_i(n); n \geq 1)$  are independent r.v.s distributed as  $\zeta'_n$ , and independent of  $(A_0, \dots, A_{i-1})$ . Observe that  $(\mathcal{S}_1)$  holds thanks to Theorem 2.1. Next, we let  $i$  be any positive integer, we assume  $(\mathcal{S}_i)$  and we prove that  $(\mathcal{S}_{i+1})$  holds. Elementary calculus shows that  $\zeta_n$  has a linear-fractional distribution, so that  $\zeta'_n$  has a geometric distribution. We are now reasoning conditionally given  $A_i = h$ . By  $(\mathcal{S}_i)$  and the definition of  $A_i$ , we get that conditional on  $A_i = h$ :

- the r.v.s  $(D_i(n); n > h)$  are independent r.v.s distributed as  $\zeta'_n$ , and independent of  $(A_0, \dots, A_{i-1})$ ,
- $D_i(h)$  has the law of  $\zeta'_h$  conditional on  $\zeta'_h \neq 0$ , and it is independent of  $(D_i(n); n > h)$  and  $(A_0, \dots, A_{i-1})$ ,
- $D_i(n) = 0$  for all  $n < h$ .

Let us apply the transition probability defined in the theorem. First,  $D_{i+1}(n) = D_i(n)$  for all  $n > h$ , so the r.v.s  $(D_{i+1}(n); n > h)$  are independent r.v.s distributed as  $\zeta'_n$ , and independent of  $(A_0, \dots, A_{i-1})$ . Second,  $D_{i+1}(h) = D_i(h) - 1$  has the law of  $\zeta'_h - 1$  conditional on  $\zeta'_h \neq 0$ , which is the law of  $\zeta'_h$  because  $\zeta'_h$  is geometrically distributed, and it is independent of  $(D_i(n); n > h)$  and  $(A_0, \dots, A_{i-1})$ . Third, the r.v.s  $(D_{i+1}(n); n < h)$  are new independent r.v.s distributed as  $\zeta'_n$ . As a consequence, conditional on  $A_i = h$ , the r.v.s  $(D_{i+1}(n); n \geq 1)$  are independent r.v.s distributed as  $\zeta'_n$ , and independent of  $(A_0, \dots, A_{i-1})$ . Integrating over  $h$  yields  $(\mathcal{S}_{i+1})$ .

We deduce from  $(\mathcal{S}_i)$  that  $A_i$  is independent of  $(A_0, \dots, A_{i-1})$  and is distributed as  $A_1$ . The computation of the law of  $A_1$  stems from well-known formulae involving linear-fractional BGW processes (see [2]), namely,

$$f_n(s) = 1 - \frac{(1 - (a/b))(1 - s)}{m^{-n}(s - (a/b)) + 1 - s},$$

if  $a \neq b$ , and

$$f_n(s) = \frac{na - (na + a - 1)s}{1 - a + na - nas},$$

when  $a = b$ . Indeed, it is then straightforward to compute

$$f'_n(0) = m^{-n} \frac{(b - a)^2}{(b - am^{-n})^2} \quad \text{and} \quad 1 - f_n(0) = \frac{b - a}{b - am^{-n}},$$

if  $a \neq b$ , whereas

$$f'_n(0) = \frac{(1 - a)^2}{(1 - a + na)^2} \quad \text{and} \quad 1 - f_n(0) = \frac{1 - a}{1 - a + na},$$

when  $a = b$ . Thanks to Theorem 2.1, the ratio of these quantities is  $\mathbb{P}(A_1 > n)$ . □

5.2. *Disintegration of the quasi-stationary distribution.* In this subsection, we assume that  $f'(1) < 1$  (subcritical case). Then it is well known [2] that there is a probability  $(\alpha_k)_{k \geq 1}$  with generating function, say  $a$ ,

$$a(s) = \sum_{k \geq 1} \alpha_k s^k, \quad s \in [0, 1],$$

such that

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = k \mid Z_n \geq 1) = \alpha_k, \quad k \geq 1.$$

This distribution is known as the *Yaglom limit*. It is a quasi-stationary distribution, in the sense that

$$\sum_{k \geq 1} \alpha_k \mathbb{P}_k(Z_1 = j \mid Z_1 \neq 0) = \alpha_j, \quad j \geq 1.$$

Set

$$U := \min\{i \geq 1 : A_i = +\infty\}.$$

Then for any  $i \leq U < j$ , the coalescence time between individuals  $(0, i)$  and  $(0, j)$  is  $\max\{A_k : i \leq k < j\} = +\infty$ , so that  $(0, i)$  and  $(0, j)$  do not have a common ancestor. Now set

$$V := \max\{A_k : 1 \leq k < U\},$$

the coalescence time of the subpopulation  $\{(0, i) : 1 \leq i \leq U\}$  (where it is understood that  $\max \emptyset = 0$ ), that is,  $-V$  is the generation of the most recent common ancestor of this subpopulation. We provide the joint law of  $(U, V)$  in the next proposition.

PROPOSITION 5.2. *The law of  $V$  is given by*

$$\mathbb{P}(V \leq n) = \alpha_1 \frac{1 - f_n(0)}{f'_n(0)} = \frac{\alpha_1}{\mathbb{P}(Z_n = 1 \mid Z_n \neq 0)}, \quad n \geq 0.$$

*Conditional on  $V = n \geq 1$ ,  $U$  has the law of  $Z_n$  conditional on  $\zeta_n \geq 2$ . In addition,  $U$  follows Yaglom’s quasi-stationary distribution, which entails the following disintegration formula:*

$$a(s) = \mathbb{P}(V = 0)s + \sum_{n \geq 1} \mathbb{P}(V = n)\mathbb{E}(s^{Z_n} \mid \zeta_n \geq 2), \quad s \in [0, 1].$$

REMARK 7. In the linear-fractional case with parameters  $(a, b)$  ( $a > b$ , subcritical case), the Yaglom quasi-stationary distribution is known to be geometric with failure probability  $b/a$ . Thanks to Proposition 5.1, the branch lengths are i.i.d. and are infinite with probability  $1 - (b/a)$ . Then thanks to the previous proposition, the quasi-stationary size  $U$  is the first  $i$  such that  $A_i$  is infinite, which indeed follows the aforementioned geometric law.

PROOF OF PROPOSITION 5.2. From the transition probabilities of the Markov chain  $(D_i; i \geq 1)$  given in Theorem 2.1, we deduce that

$$V = \max\{n : D_1(n) \neq 0\},$$

so that, thanks to Theorem 2.1,

$$\mathbb{P}(V \leq n) = \prod_{k \geq n+1} \mathbb{P}(\zeta'_k = 0) = \frac{\prod_{k \geq 1} \mathbb{P}(\zeta'_k = 0)}{\prod_{k=1}^n \mathbb{P}(\zeta'_k = 0)} = \frac{\mathbb{P}(A_1 = +\infty)}{\mathbb{P}(A_1 > n)}.$$

Now, since

$$\mathbb{P}(A_1 > n) = \frac{f'_n(0)}{1 - f_n(0)} = \mathbb{P}(Z_n = 1 \mid Z_n \neq 0),$$

and because this last quantity converges to  $\mathbb{P}(A_1 = +\infty) = \alpha_1$  as  $n \rightarrow \infty$ , we get the result for the law of  $V$ .

Recall that  $(-n, a_1(n))$  is the ancestor, at generation  $-n$ , of  $(0, 1)$ , so that the total number of descendants  $\Upsilon_n := Z^{(-n, a_1(n))}(n)$  of  $(-n, a_1(n))$  at generation 0 has the law of  $Z_n$  conditional on  $Z_n \neq 0$ . Now, since no individual  $(0, j)$  with  $j > U$  has a common ancestor with  $(0, 1)$ , we have the inequality  $\Upsilon_n \leq U$ . On the other hand,  $(0, U)$  and  $(0, 1)$  do have a common ancestor [all coalescence times  $(A_i; 1 \leq i \leq U - 1)$  are finite], so that there is  $n$  such that  $\Upsilon_n = U$ . Since the sequence  $(\Upsilon_n)$  is obviously nondecreasing, it is stationary at  $U$ , that is,

$$\lim_{n \rightarrow \infty} \Upsilon_n = U \quad \text{a.s.}$$

Since  $\Upsilon_n$  has the law of  $Z_n$  conditional on  $Z_n \neq 0$ ,  $U$  follows the Yaglom quasi-stationary distribution.

Actually, since  $-V$  is the generation of the most recent common ancestor of  $(0, 1)$  and  $(0, U)$ , we have the following equality:

$$U = \Upsilon_V.$$

Now recall that  $V = \max\{n : D_1(n) \neq 0\}$ . By definition of  $D_1$ , we can write  $\{V = n\} = E_n \cap F_n$ , where  $E_n := \{D_1(k) = 0, \forall k > n\}$  and

$$F_n := \{D_1(n) \geq 1\} = \{\#D(n, 1) \geq 2\}.$$

Now observe that  $E_n$  is independent of all events of the form  $F_n \cap \{\Upsilon_n = k\}$  (it concerns the future of parallel branches), so that  $\mathbb{P}(\Upsilon_n = k \mid V = n) = \mathbb{P}(\Upsilon_n = k \mid$

$F_n$ ). In other words,  $\Upsilon_n$  conditional on  $V = n$  has the law of  $Z'_n$  conditional on  $\zeta_n \geq 2$ , where  $Z'_n$  is  $Z_n$  conditional on  $Z_n \neq 0$ . Since  $\{Z_n \neq 0\} = \{\zeta_n \geq 1\}$ , we finally get that conditional on  $V = n$ ,  $U$  has the law of  $Z_n$  conditional on  $\zeta_n \geq 2$ .  $\square$

**Acknowledgments.** A. Lambert wishes to thank Julien Berestycki and Olivier Hénard for some interesting discussions on the topic of this paper. A. Lambert and L. Popovic thank the referees for suggestions that improved the exposition of the paper.

## REFERENCES

- [1] ALDOUS, D. and POPOVIC, L. (2005). A critical branching process model for biodiversity. *Adv. in Appl. Probab.* **37** 1094–1115. [MR2193998](#)
- [2] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes. Die Grundlehren der Mathematischen Wissenschaften, Band 196*. Springer, New York. [MR0373040](#)
- [3] BENNIES, J. and KERSTING, G. (2000). A random walk approach to Galton–Watson trees. *J. Theoret. Probab.* **13** 777–803. [MR1785529](#)
- [4] BERTOIN, J. and LE GALL, J.-F. (2000). The Bolthausen–Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields* **117** 249–266. [MR1771663](#)
- [5] BERTOIN, J. and LE GALL, J.-F. (2006). Stochastic flows associated to coalescent processes. III. Limit theorems. *Illinois J. Math.* **50** 147–181 (electronic). [MR2247827](#)
- [6] BIRKNER, M., BLATH, J., CAPALDO, M., ETHERIDGE, A., MÖHLE, M., SCHWEINSBERG, J. and WAKOLBINGER, A. (2005). Alpha-stable branching and beta-coalescents. *Electron. J. Probab.* **10** 303–325 (electronic). [MR2120246](#)
- [7] CHAUVIN, B., ROUAULT, A. and WAKOLBINGER, A. (1991). Growing conditioned trees. *Stochastic Process. Appl.* **39** 117–130. [MR1135089](#)
- [8] DUQUESNE, T. and LE GALL, J.-F. (2002). Random trees, Lévy processes and spatial branching processes. *Astérisque* **281** vi+147. [MR1954248](#)
- [9] EVANS, S. N. (1993). Two representations of a conditioned superprocess. *Proc. Roy. Soc. Edinburgh Sect. A* **123** 959–971. [MR1249698](#)
- [10] FLEISCHMANN, K. and SIEGMUND-SCHULTZE, R. (1977). The structure of reduced critical Galton–Watson processes. *Math. Nachr.* **79** 233–241. [MR0461689](#)
- [11] GEIGER, J. (1996). Size-biased and conditioned random splitting trees. *Stochastic Process. Appl.* **65** 187–207. [MR1425355](#)
- [12] GEIGER, J. (1999). Elementary new proofs of classical limit theorems for Galton–Watson processes. *J. Appl. Probab.* **36** 301–309. [MR1724856](#)
- [13] GERNHARD, T. (2008). The conditioned reconstructed process. *J. Theoret. Biol.* **253** 769–778.
- [14] GRIMVALL, A. (1974). On the convergence of sequences of branching processes. *Ann. Probab.* **2** 1027–1045. [MR0362529](#)
- [15] KALLENBERG, O. (1977). Stability of critical cluster fields. *Math. Nachr.* **77** 7–43. [MR0443078](#)
- [16] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. [MR1876169](#)
- [17] KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. [MR0671034](#)
- [18] LAMBERT, A. (2002). The genealogy of continuous-state branching processes with immigration. *Probab. Theory Related Fields* **122** 42–70. [MR1883717](#)

- [19] LAMBERT, A. (2003). Coalescence times for the branching process. *Adv. in Appl. Probab.* **35** 1071–1089. [MR2014270](#)
- [20] LAMBERT, A. (2007). Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Electron. J. Probab.* **12** 420–446. [MR2299923](#)
- [21] LAMBERT, A. (2008). Population dynamics and random genealogies. *Stoch. Models* **24** 45–163. [MR2466449](#)
- [22] LAMBERT, A. (2009). The allelic partition for coalescent point processes. *Markov Process. Related Fields* **15** 359–386. [MR2554367](#)
- [23] LAMBERT, A. (2010). The contour of splitting trees is a Lévy process. *Ann. Probab.* **38** 348–395. [MR2599603](#)
- [24] LE GALL, J.-F. and LE JAN, Y. (1998). Branching processes in Lévy processes: The exploration process. *Ann. Probab.* **26** 213–252. [MR1617047](#)
- [25] LYONS, R., PEMANTLE, R. and PERES, Y. (1995). Conceptual proofs of  $L \log L$  criteria for mean behavior of branching processes. *Ann. Probab.* **23** 1125–1138.
- [26] POPOVIC, L. (2004). Asymptotic genealogy of a critical branching process. *Ann. Appl. Probab.* **14** 2120–2148. [MR2100386](#)
- [27] RANNALA, B. (1997). On the genealogy of a rare allele. *Theoret. Popul. Biol.* **52** 216–223.
- [28] STADLER, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *J. Theoret. Biol.* **261** 58–66.

ÉQUIPE PROBABILITÉS, STATISTIQUES  
& BIOLOGIE  
LABORATOIRE DE PROBABILITÉS  
& MODÈLES ALÉATOIRES  
UPMC UNIV PARIS 06  
CASE COURRIER 188  
4, PLACE JUSSIEU  
75252 PARIS CEDEX 05  
FRANCE

DEPARTMENT OF MATHEMATICS AND STATISTICS  
CONCORDIA UNIVERSITY  
1455 DE MAISONNEUVE BLVD WEST  
MONTREAL, H3G 1M8  
CANADA