Reversing the cut tree of the Brownian continuum random tree

Nicolas Broutin * Minmin Wang †

August 13, 2014

Abstract

Consider the logging process of the Brownian continuum random tree (CRT) $\mathcal{T}$ using a Poisson point process of cuts on its skeleton [Aldous and Pitman, *Ann. Probab.*, vol. 26, pp. 1703–1726, 1998]. Then, the cut tree introduced by Bertoin and Miermont describes the genealogy of the fragmentation of $\mathcal{T}$ into connected components [*Ann. Appl. Probab.*, vol. 23, pp. 1469–1493, 2013]. This cut tree $\mathrm{cut}(\mathcal{T})$ is distributed as another Brownian CRT, and is a function of the original tree $\mathcal{T}$ and of the randomness in the logging process. We are interested in reversing the transformation of $\mathcal{T}$ into $\mathrm{cut}(\mathcal{T})$: we define a *shuffling* operation, which given a Brownian CRT $\mathcal{H}$, yields another one $\mathrm{shuff}(\mathcal{H})$ distributed in such a way that $(\mathcal{T}, \mathrm{cut}(\mathcal{T}))$ and $(\mathrm{shuff}(\mathcal{H}), \mathcal{H})$ have the same distribution.

1 Introduction

Let $\mathcal{T}$ be Aldous' Brownian continuum random tree (CRT) [2]. To the logging process of $\mathcal{T}$ introduced in Aldous and Pitman [4], one can associate another continuum random tree $\mathrm{cut}(\mathcal{T})$, which describes the genealogical structure of this fragmentation process (see [10] and [13]). Moreover, for a Brownian CRT, this associated tree $\mathrm{cut}(\mathcal{T})$ is also distributed as a Brownian CRT. One of the main questions [Miermont, Pers. Comm.] then is whether the transformation from $\mathcal{T}$ to the genealogy of the fragmentation $\mathrm{cut}(\mathcal{T})$ is "reversible". Of course, some information has been lost about the initial tree $\mathcal{T}$, and one must first understand whether it is possible to resample this information, and then study the possibility of the construction of a tree $\mathcal{T}'$ that is distributed like $\mathcal{T}$, conditional on $\mathrm{cut}(\mathcal{T})$.

Let $\mathcal{H}$ be another Brownian CRT (that should be informally thought of as $\mathrm{cut}(\mathcal{T})$), we define below a continuous tree $\mathrm{shuff}(\mathcal{H})$, which is random given $\mathcal{H}$, and such that the following identity in distribution holds:

$$\big(\mathrm{shuff}(\mathcal{H}), \mathcal{H}\big) \stackrel{d}{=} \big(\mathcal{T}, \mathrm{cut}(\mathcal{T})\big). \tag{1}$$

The construction of the tree $\mathrm{shuff}(\mathcal{H})$ from $\mathcal{H}$ is the main objective of the present document, and can be seen as follows. Let $\mathrm{Br}(\mathcal{H})$ denote the set of branch points of $\mathcal{H}$. Start by assigning independently to every branch point $x \in \mathrm{Br}(\mathcal{H})$ a random point A_x sampled using the mass measure ν restricted to $\mathrm{Sub}(\mathcal{H}, x)$, the subtree of $\mathcal{H}$ above x. For each such x, the choice of A_x induces a choice among the subtrees of $\mathrm{Sub}(\mathcal{H}, x)$ rooted at x: Let the *fringe* $\mathrm{Fr}(\mathcal{H}, x, A_x)$ be the subset of points $y \in \mathrm{Sub}(\mathcal{H}, x)$ for which the closest common ancestor of y and A_x is $y \wedge A_x = x$. Then, informally $\mathrm{shuff}(\mathcal{H})$ is obtained by detaching $\mathrm{Fr}(\mathcal{H}, x, A_x)$ and reattaching it at A_x, for every branch point x of $\mathcal{H}$ (see Figure 1); the points of the skeleton that are not branch points are not used. It is a priori unclear whether this definition makes sense, let alone that the resulting metric space is a real tree or that it has the correct distribution.

*Inria Paris–Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay - France. Email: nicolas.broutin@intria.fr
†Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris - France. Email: wangminmin03@gmail.com

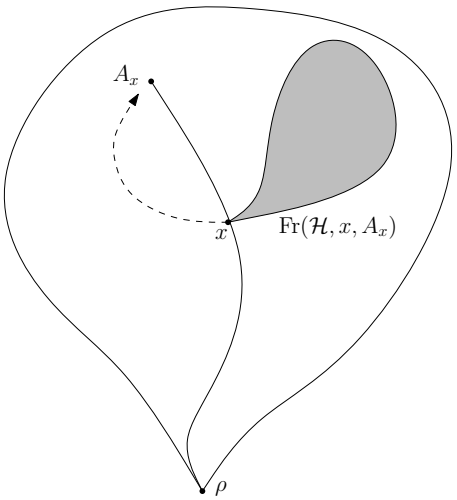

Figure 1: *The surgical operation on the tree $\mathcal{H}$, rooted at ρ, for a single branch point x with corresponding attach point A_x.*

It indeed seems that we discard from $\mathcal{H}$ all the length by leaving the skeleton behind. The remainder of the document is devoted to making this construction rigorous, and to prove that the tree $\mathrm{shuff}(\mathcal{H})$ satisfies (1).

The continuous problem at hand is connected to a rather large body of work on the destruction of random (discrete) trees by sampling of random nodes or edges initiated by Meir and Moon [24]. There is no significant difference between sampling nodes or edges, and we present here a version that samples nodes and proceeds as follows: sample a random node in some rooted tree (random or not), discard the portion that is now disconnected from the root, and keep going until the root is finally picked (the process then stops). The main question addressed by Meir and Moon [24] and many of the researchers after them was about the number of steps, or cuts, that are needed for the process to terminate. This problem has been considered for a number of classical models of trees including random binary search trees [19, 20], random recursive trees [5, 8, 14, 21] and the family trees of Galton–Watson processes conditioned on the total progeny [1, 16, 22, 25]. Janson [22] was the first to realize by moment calculations that, when the tree is a Galton–Watson tree, there should be nice constructions of the limit random variables directly in terms of continuous cutting of the Brownian CRT. In some sense, the continuous cutting alluded in [22] is just a version of the logging of the CRT in which only the cuts affecting the size of the connected component containing the root are retained. The constructions in [1], [7] and [10] all encode the "number of cuts" affecting the connected components containing some points as the total length of some distinguished subtrees. The construction of the genealogy as a compact tree is due to Bertoin and Miermont [10].

Let us now describe our approach to the definition of $\mathrm{shuff}(\mathcal{H})$. The idea is to construct it by defining an order in which the fringes should be sent to their new attach points. This ordering yields a tree-valued Markov chain, and we formally define $\mathrm{shuff}(\mathcal{H})$ as its almost sure limit. More precisely, a construction of the first element of this Markov chain appears in [1] and has been formally justified in [13]: there the subtrees to be reattached are only those lying along the path between the root and a distinguished random leaf U_1. In the following, we refer to this transformation as the *one-path reversal*. The Markov chain we have in mind consists in iteratively reattaching the subtrees lying on the paths to an i.i.d. sequence of leaves $(U_i)_{i\geq 1}$ in $\mathcal{H}$, that we later refer to as the *i-paths reversals*, or *i-reversals* for short. However, not any such sequence $(U_i)_{i\geq 1}$ would do. Indeed, although it is very close to the one we are after, the one-path reversal enforces that the subtrees to be detached are precisely $\mathrm{Fr}(\mathcal{H}, x, U_1)$, for the branch points x on the path to U_1; so in particular, they are only defined in terms of U_1, and the choices A_x are

then somewhat conditioned on being consistent with the constraint that

$$\mathrm{Fr}(\mathcal{H}, x, A_x) = \mathrm{Fr}(\mathcal{H}, x, U_1). \tag{2}$$

It follows that if we want to use the results of [1, 13], then the sequence $(U_i)_{i\geq 1}$ must be constructed from $(A_x, x \in \mathrm{Br}(\mathcal{H}))$ in such a way that, for all the branch points x on the path to U_i, the constraint in (2) is satisfied with U_i instead of U_1 in the right-hand side.

Plan of the paper. The route we use here to define $\mathrm{shuff}(\mathcal{H})$ relies on a careful understanding of the cutting procedure and of the genealogy induced by finitely many random points only. In Section 2, we introduce the relevant background on cut trees, and their reversals. We also prove a few results that have not appeared elsewhere. Section 3 is devoted to proving that the sequence of k-paths reversals converges as $k \to \infty$ in the sense of Gromov–Prokhorov. Up to this point, the shuffle operation is therefore justified as a refining sequence of k-reversals. The direct construction presented above is then justified in Section 4 by proving that one can construct a sequence of leaves such that the shuffle tree corresponds to the limit of the k-reversals with respect to this sequence of leaves. Some auxiliary results about the Brownian CRT for which we did not find a reference are proved in Appendix A.

2 Preliminaries on cut trees and shuffle trees

In this section, we recall the previous results in [13] on the cut trees and the shuffle trees of the Brownian continuum random tree.

2.1 Notations and background on continuum random trees

We only give here a short overview, the interested reader may consult [3], [23], or [15] for more details.

A *real tree* is a geodesic metric space without loops. The real trees we are interested in are compact. A *continuum random tree* T is a random (rooted) real tree equipped with a probability measure, often referred to as the *mass measure* or the uniform measure. The *Brownian continuum random tree* is a special continuum random tree that has been introduced by Aldous [2] as the scaling limit of uniformly random trees. One way to define the Brownian CRT starts from a standard normalized Brownian excursion of unit length $e = (e_s, 0 \leq s \leq 1)$. For any $s, t \in [0, 1]$, let

$$\frac{1}{2} d(s,t) := e_s + e_t - 2 \inf_{u \in [s,t]} e_u, \tag{3}$$

and define $s \sim t$ if $d(s,t) = 0$. Then d induces a metric on the quotient space $[0,1]/\!\sim$. Moreover, this metric space is a real tree: it is the Brownian CRT, which we denote by $(\mathcal{T}, d_{\mathcal{T}})$ in the following. A Brownian CRT $\mathcal{T}$ also comes with a *mass measure* $\mu_{\mathcal{T}}$, which is the push-forward of Lebesgue measure by the canonical projection $p : [0,1] \to \mathcal{T}$. A point that is sampled according to $\mu_{\mathcal{T}}$ is usually called here a $\mu_{\mathcal{T}}$-point. The Brownian CRT $\mathcal{T}$ is rooted at the point $p(0)$. For $a > 0$, let $e^{(a)} = (e^{(a)}_s, 0 \leq s \leq a)$ denote the Brownian excursion of length a. We can associate with $e^{(a)}$ a random real tree, denoted by $\mathcal{T}^{(a)}$, by replacing e with $e^{(a)}$ in (3). If $s > 0$, we denote by $s\mathcal{T}$ the metric space in which the distance is $s d_{\mathcal{T}}$. Then, the Brownian scaling implies that (see also Appendix A)

$$\mathcal{T}^{(a)} \overset{d}{=} \sqrt{a}\mathcal{T}.$$

And clearly, the mass measure of $\mathcal{T}^{(a)}$, which is the push-forward of Lebesgue measure on $[0, a]$, has total mass a. In what follows, we sometimes refer to $\mathcal{T}$ as the standard Brownian CRT.

For $u, v \in \mathcal{T}$, we denote by $[\![u, v]\!]$ and $]\!]u, v[\![$ the closed and open paths between u and v in $\mathcal{T}$, respectively. For $u \in \mathcal{T}$, the degree of u in $\mathcal{T}$, denoted by $\deg(u, \mathcal{T})$, is the number of connected components of $\mathcal{T} \setminus \{u\}$. We also denote by

$$\mathrm{Lf}(\mathcal{T}) = \{u \in \mathcal{T} : \deg(u, \mathcal{T}) = 1\} \quad \text{and} \quad \mathrm{Br}(\mathcal{T}) = \{u \in \mathcal{H} : \deg(u, \mathcal{T}) \geq 3\}$$

the set of the *leaves* and the set of *branch points* of $\mathcal{T}$, respectively. Almost surely, these two sets are everywhere dense in $\mathcal{T}$, though $\mathrm{Lf}(\mathcal{T})$ is uncountable and $\mathrm{Br}(\mathcal{T})$ countable. The skeleton of $\mathcal{T}$ is the complement of $\mathrm{Lf}(\mathcal{T})$ in $\mathcal{T}$, denoted by $\mathrm{Sk}(\mathcal{T})$. The skeleton is the union

$$\mathrm{Sk}(\mathcal{T}) = \bigcup_{u,v \in \mathcal{T}}]\!]u, v[\![.$$

If ρ is the root of $\mathcal{T}$, for $u \in \mathcal{T}$, the subtree above u, denoted by $\mathrm{Sub}(\mathcal{T}, u)$, is defined to be the subset $\{v \in \mathcal{T} : u \in [\![\rho, v]\!]\}$. If $v \in \mathrm{Sub}(\mathcal{T}, u)$ is distinct from u, we denote by $\mathrm{Fr}(\mathcal{T}, u, v)$ the fringe tree hung from $[\![u, v]\!]$ which is the set $\{w \in \mathrm{Sub}(\mathcal{T}, u) : [\![w, u]\!] \cap [\![u, v]\!] = \{u\}\}$. It is nontrivial only if $u \in \mathrm{Br}(\mathcal{T})$. There also exists a unique σ-finite measure ℓ concentrated on $\mathrm{Sk}(\mathcal{T})$ such that for any two points $u, v \in \mathcal{T}$ we have $\ell([\![u, v]\!]) = d_{\mathrm{T}}(u, v)$; ℓ is called the *length measure*. If ρ denotes the root of $\mathcal{T}$ and $v_1, \cdots, v_k$ are k points of $\mathcal{T}$, we write

$$\mathrm{Span}(\mathrm{T}; v_1, \cdots, v_k) = \bigcup_{1 \leq i \leq k} [\![\rho, v_i]\!]$$

for the subtree of $\mathcal{T}$ spanning the root ρ and $\{v_i, 1 \leq i \leq k\}$.

The state space of interest is the set of metric spaces that are pointed, that is with a distinguished point that we call the root and equipped with a probability measure. More precisely, it is the set of equivalence classes induced by measure-preserving isometries (on the support of the probability measure). When equipped with the Gromov–Prokhorov (GP) distance, this yields a Polish space. Convergence in the GP topology is equivalent to convergence in distribution of the matrices whose entries are distances between the pairs of points sampled from the probability distribution μ. This is discussed at length in [13], and we also refer the reader to [18] and [17] for more information.

2.2 The cutting procedure on a Brownian CRT

Let $\mathcal{T}$ be a Brownian CRT to be cut down. Now let $\mathcal{P}$ be a Poisson point process of intensity measure $dt \otimes \ell(dx)$ on $\mathbb{R}_+ \times \mathcal{T}$. Every point $(t, x) \in \mathcal{P}$ is seen as a cut on $\mathcal{T}$ at location x which arrives at time t. Then $\mathcal{P}$ defines a Poisson rain of cuts that split $\mathcal{T}$ into smaller and smaller connected components as time goes. More precisely, let $(V_i)_{i \geq 1}$ be a sequence of independent points sampled according to μ, then for each $t \geq 0$, $\mathcal{P}$ induces a nested process of exchangeable partitions of $\mathbb{N}$ in the following way. For each $t \geq 0$, the blocks are the equivalence classes of the relation $\sim_t$ defined by

$$i \sim_t j \quad \text{if and only if} \quad \mathcal{P} \cap ([0, t] \times [\![V_i, V_j]\!]) = \varnothing.$$

Let $\mathcal{T}_i(t)$ be the set of those points in $\mathcal{T}$ which are still connected to V_i at time t, that is

$$\mathcal{T}_i(t) := \{u \in \mathcal{T} : \mathcal{P} \cap ([0, t] \times [\![V_i, u]\!]) = \varnothing\}.$$

Then it is easy to see that $\mathcal{T}_i(t)$ is a connected subspace of $\mathcal{T}$, that is, a subtree of $\mathcal{T}$. Furthermore, we have $\mathcal{T}_i(t) \subseteq \mathcal{T}_i(s)$ if $s \leq t$, and $\cap_{t \geq 0} \mathcal{T}_i(t) = \{V_i\}$ almost surely, since with probability one the atoms of $\mathcal{P}$ are everywhere dense in $\mathcal{T}$ and V_i is not among these atoms.

2.3 The k-cut tree

The main point of the definition of a cut tree is to obtain a representation of the genealogy of the fragmentation induced by $\mathcal{P}$ as a compact real tree. A first step consists in focusing on the genealogy of the fragmentation induced on $[k] = \{1, 2, \dots, k\}$, for some $k \geq 1$. So at this point, we only keep track of the evolution of the connected components containing the points $V_1, V_2, \cdots, V_k$ and ignore all the other ones.

For each $t \geq 0$, let us write $\pi_k(t)$ for the partition of $[k]$ induced by $\sim_t$. Then for any $t \geq s$, $\pi_k(t)$ is a refinement of $\pi_k(s)$. We encode the family $(\pi_k(t))_{t\geq 0}$ by a rooted tree S_k with k leaves. Each equivalence class induced on $[k]$ by some $\sim_t$, $t \geq 0$ is represented by a *node* of S_k. It is also convenient to add an additional node r, which we see as the root of S_k. From the root r there is a unique edge, which connects r to the node labeled by $[k] := \{1, 2, \cdots, k\}$. Let $t_{[k]} := \sup\{t \geq 0 : \pi_k(t) \neq \{[k]\}\}$ be the time when $[k]$ disappears from $(\pi_k(t))_{t\geq 0}$. Note that $t_{[k]}$ is the first moment when there is some point $x_{[k]}$ of the subtree of $\mathcal{T}$ spanning $\{V_1, V_2, \dots, V_k\}$,

$$\bigcup_{1\leq i,j\leq k} [\![V_i, V_j]\!]$$

such that $(t_{[k]}, x_{[k]}) \in \mathcal{P}$. Observe that almost surely $x_{[k]}$ has degree two in $\mathcal{T}$, so that $\pi_k(t_{[k]})$ consists of only two blocks E_1 and E_2 with probability one. This is represented in S_k by the fact that the node labelled $[k]$ has two children, labeled respectively by E_1 and E_2. One then proceeds recursively to define the subtrees induced on the set of leaves in E_1 and E_2, respectively. We obtain a binary tree on k leaves labelled by $\{1\}, \cdots, \{k\}$. (See Figure 2).

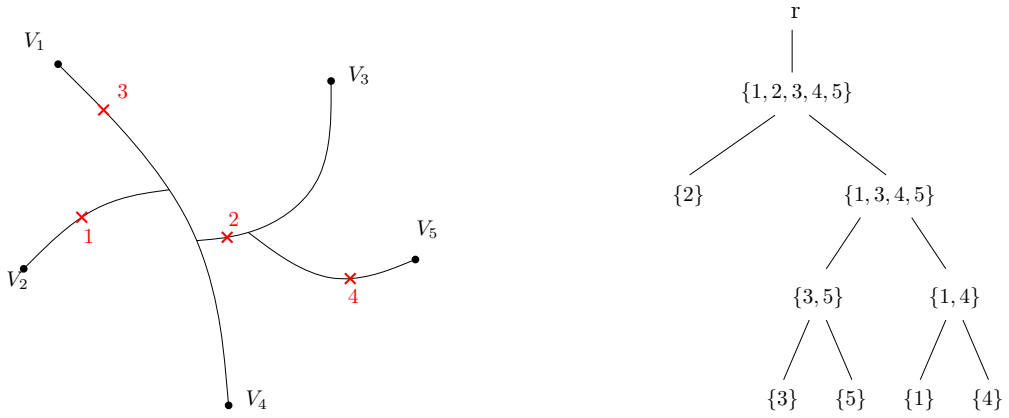

Figure 2: *On the left, the subtree of $\mathcal{T}$ spanning the leaves $V_1, V_2, \dots, V_5$. The cuts falling on it are represented by the crosses, and the index next to them is the time at which they appear. On the right, the corresponding tree S_k.*

We now endow S_k with a distance d_{S_k}, or to be more precise we define a binary real tree that has the same tree structure as S_k. For this, we set for $1 \leq i \leq k$ and $t \in [0, \infty]$,

$$L_i(t) := \int_0^t \mu\big(\mathcal{T}_i(s)\big)ds.$$

Then $L_i(\infty)$ is finite almost surely (this is shown for instance in [1]). For every $i \in [k]$, we want to identify the unique path of S_k from the root $[k]$ to the leaf $\{i\}$ with the finite interval $\{L_i(t) : t \in [0, \infty]\}$ such that if a node is labelled by E for some $E \subseteq [k]$, then it is at distance $L_i(t_E)$ from the root, where $t_E = \sup\{t \geq 0 : E \in \pi_k(t)\}$. Doing so does not cause any ambiguity since if $i, j \in E$ then $\mathcal{T}_j(s) = \mathcal{T}_i(s)$ for any $s \leq t_E$. So we obtain a compact real tree which consists of k paths of respective lengths $L_i(\infty), 1 \leq i \leq k$. By slightly abusing the notation, we still write S_k for the real tree (S_k, d_{S_k}).

In other words, if we write $E_i(0) = r, E_i(1) = [k], E_i(2), \dots, E_i(h_i) = \{i\}$ for the sequence of nodes on the path from the root to $\{i\}$ in S_k, the real tree (S_k, d_{S_k}) is the tree S_k in which the edges have been replaced by the $2k-1$ intervals of lengths $L_i(t_{E_i(h+1)}) - L_i(t_{E_i(h)})$, for $0 \le h < h_i$ and $1 \le i \le k$.

We now move on to the definition of the k-cut tree. The real tree S_k provides the *backbone* of the k-cut tree. We define the k-cut tree $\mathrm{cut}(\mathcal{T}, V_1, \cdots, V_k)$ as the real tree obtained by grafting on the backbone S_k the subtrees discarded during the cutting procedure. Let $\mathcal{C}^k$ be the set of those $t \ge 0$ for which

$$\mu\Big(\bigcup_{1\le i\le k} \mathcal{T}_i(t)\Big) < \mu\Big(\bigcup_{1\le i\le k} \mathcal{T}_i(t-)\Big),$$

Note that $\mathcal{C}^k$ is almost surely countable. If $t \in \mathcal{C}^k$, let i_t be the smallest element of $[k]$ such that $\Delta\mathcal{T}_{i_t}(t) := \mathcal{T}_{i_t}(t-) \setminus \bigcup_{1\le j\le k} \mathcal{T}_j(t) \neq \varnothing$. Because of the holes left by the previous cuts, $\Delta\mathcal{T}_{i_t}(t)$ is a connected but not complete subspace in $\mathcal{T}$. We let Δ_t^k be the completion of $\Delta\mathcal{T}_{i_t}(t)$. Almost surely there exists a unique $x \in \mathcal{T}$ such that $(t, x) \in \mathcal{P}$. Note that $x \in \Delta\mathcal{T}_{i_t}(t)$. We denote by $x' \in \Delta_t^k$ the image of x via the canonical injection from $\Delta\mathcal{T}_{i_t}(t)$ to Δ_t^k. We think of Δ_t^k as rooted at x'. Then for each $t \in \mathcal{C}^k$, we graft Δ_t^k by its root on the path in S_k connecting the root to the leaf $\{i_t\}$ at distance $L_{i_t}(t)$ from the root. We denote by $\mathcal{G}_k = \mathrm{cut}(\mathcal{T}, V_1, \cdots, V_k)$ the obtained metric space. The tree $\mathcal{G}_k$ also bears a mass measure which is inherited from that of $\mathcal{T}$; the set of the points which have been added (either in the backbone S_k or due to completion) is assigned mass 0. The new mass is still denoted by μ. An alternative way to define $\mathcal{G}_k$ (which is the way we have used in [13]) is to graft $\Delta\mathcal{T}_{i_t}(t)$ (rather than Δ_t^k) on S_k, and then to complete the metric space. One easily checks that these two definitions coincide.

Remark. There is a number of different mass measures that we need to consider here. In order to clarify the discussion and to keep the notation under control, we have decided to keep using the same name for the mass measure when only a set of measure zero was modified by the transformation either by removal of countably many points, by (countable) completion, or by the addition of a backbone. For instance, we think of the tree $\mathcal{G}_k$ as still carrying the mass measure μ of $\mathcal{T}$.

Proposition 1 (Distribution of the k-cut tree)**.** *If $\mathcal{T}$ is the Brownian CRT, and $(V_i)_{i\ge1}$ is a sequence of i.i.d. points of $\mathcal{T}$ with common distribution μ, then for each $k \ge 1$, we have*

$$(\mathcal{G}_k, S_k) \overset{d}{=} \big(\mathcal{T}, \mathrm{Span}(\mathcal{T}; V_1, V_2, \cdots, V_k)\big). \tag{4}$$

The case $k = 1$ corresponds to a special case of Theorem 3.2 in [13]. The general case follows from the analogous result on the discrete trees (see Section 4, Lemma 4.5 there) and the same weak convergence argument as in [13], and we omit the details.

The "complete" cut tree By construction, since the definition of $\mathcal{T}_i(t)$ does not depend on k, we have $S_k \subset S_{k+1}$ for each $k \ge 1$.

Proposition 2 (Complete cut tree, [10, 13])**.** *Let $\mathrm{cut}(\mathcal{T}) = \overline{\cup_k S_k}$ be the limit metric space of $(S_k)_{k\ge1}$. If $\mathcal{T}$ is the Brownian CRT, then almost surely, $\mathrm{cut}(\mathcal{T})$ is a compact real tree and is distributed as $\mathcal{T}$.*

The construction of $\mathcal{G}_k$ described above yields the following recurrence relation between $\mathcal{G}_{k+1}$ and $\mathcal{G}_k$ (see Figure 3). For every $k \ge 1$, the collection Δ_t^k, $t \in \mathcal{C}^k$ has full mass and a uniform point V falls with probability one in Δ_t^k, for some $t \in \mathcal{C}^k$. If we let $m_k := \mu(\Delta_t^k)$, then $\mathrm{m}_k{}^{-1/2}\Delta_t^k$ is distributed as a standard Brownian CRT. As a consequence, the 1-cut tree $\mathrm{cut}(\Delta_t^k, V)$ is well-defined.

Proposition 3 (Recurrence relation for $(\mathcal{G}_k)_{k\ge0}$)**.** *For each $k \ge 2$, let $\tau_k \in \mathcal{C}^{k-1}$ be such that $V_k \in \Delta_{\tau_k}^{k-1}$. Then $\mathcal{G}_k$ is obtained from $\mathcal{G}_{k-1}$ by replacing $\Delta_{\tau_k}^{k-1}$ with $\mathrm{cut}(\Delta_{\tau_k}^{k-1}, V_k)$.*

The cut tree $\mathrm{cut}(\mathcal{T})$ may then be seen as the limit of the k-cut trees. The notion of $\mathrm{cut}(\mathcal{T})$ here coincides with that in [8].

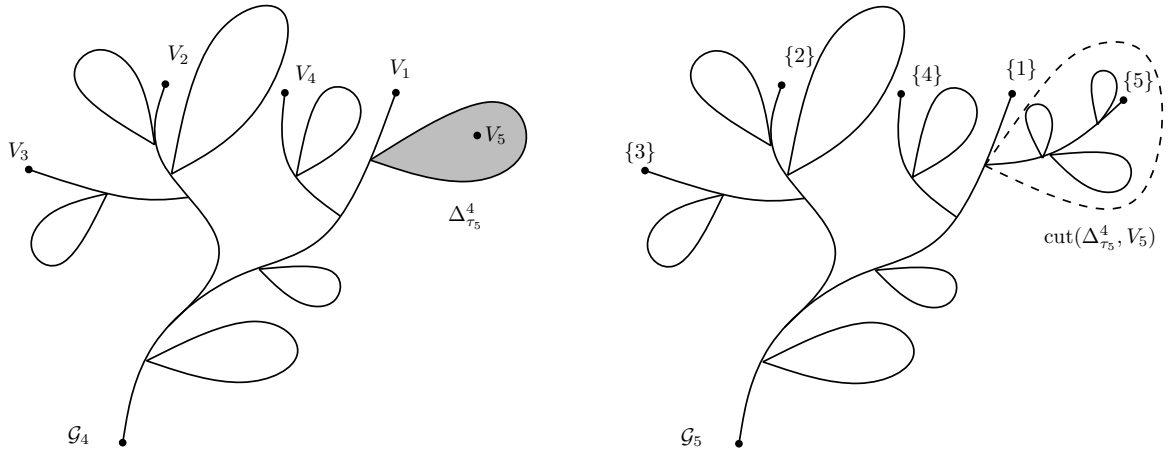

Figure 3: *The transformation from $\mathcal{G}_{k-1} = \mathrm{cut}(\mathcal{T}, V_1, \ldots, V_{k-1})$ to $\mathcal{G}_k = \mathrm{cut}(\mathcal{T}, V_1, \ldots, V_{k-1}, V_k)$ with $k = 5$: with probability one, V_k falls in some $\Delta^{k-1}_{\tau_k}$, a connected component of $\mathcal{G}_{k-1} \setminus S_{k-1}$. That subtree stopped being transformed at time τ_k, since it did not contain any of $V_1, \ldots, V_{k-1}$, and it should now be replaced by $\mathrm{cut}(\Delta^{k-1}_{\tau_k}; V_k)$.*

Proposition 4 (Convergence of k-cut trees)**.** *Let $\mathcal{T}$ be the Brownian CRT. As $k \to \infty$,*

$$\mathrm{cut}(\mathcal{T}, V_1, \cdots, V_k) \to \mathrm{cut}(\mathcal{T}), \quad \textit{almost surely}$$

in the Gromov–Hausdorff topology.

Proof. If (T, d) is a compact real tree rooted at ρ, we let $\mathrm{ht}(T) = \sup_{u \in T} d(u, \rho)$ and $\mathrm{diam}(T) = \sup_{u,v \in T} d(u, v)$ denote the respective height and diameter of T. By the triangle inequality, we have $\mathrm{ht}(T) \leq \mathrm{diam}(T) \leq 2\,\mathrm{ht}(T)$. On the one hand, we deduce from Proposition 3 that

$$\Upsilon_k := \sup_{t \in \mathcal{C}^k} \mathrm{diam}(\Delta^k_t), \quad k \geq 1$$

is a non-increasing sequence, so it converges almost surely to some random variable that we denote by Υ_∞. On the other hand, if we write δ_{H} for the Hausdorff distance (on the compact subsets of $\mathcal{G}_k$), then it follows from the construction of $\mathcal{G}_k$ that

$$\delta_{\mathrm{H}}(\mathcal{G}_k, S_k) = \sup_{t \in \mathcal{C}^k} \mathrm{ht}(\Delta^k_t).$$

Proposition 1 implies that the sequence $(\delta_{\mathrm{H}}(\mathcal{G}_k, S_k))_{k \geq 1}$ converges to 0 in distribution. Combining this with the bounds $\delta_{\mathrm{H}}(\mathcal{G}_k, S_k) \leq \Upsilon_k \leq 2\,\delta_{\mathrm{H}}(\mathcal{G}_k, S_k)$, we obtain that $\Upsilon_\infty = 0$, which then entails that $\delta_{\mathrm{H}}(\mathcal{G}_k, S_k) \to 0$ almost surely as $k \to \infty$. Since, by Proposition 2, we have $\lim_{k \to \infty} \delta_{\mathrm{H}}(\mathrm{cut}(\mathcal{T}), S_k) = 0$ a.s., the result follows. □

Distribution of a uniform path One of the main ingredients of our construction of the shuffle tree in Section 3 consists in understanding how the path between two points gets transformed as the approximation $\mathcal{G}_k$ of the cut tree $\mathrm{cut}(\mathcal{T})$ gets refined. More precisely, let ξ_1, ξ_2 be two independent μ-points of $\mathcal{T}$ and write $p :=]\!]\xi_1, \xi_2[\![$ for the open path between them in $\mathcal{T}$. Initially, in $\mathcal{G}_0 := \mathcal{T}$, p is indeed a path. But later on, as k increases, this path p gets cut into pieces each contained in some of the Δ^k_t that are grafted onto the backbone S_k. More formally, for each $k \geq 0$, there exists an injective map $\phi^\circ_k : \cup_{t \in \mathcal{C}^k} \Delta^k_t \to \mathcal{G}_k$ whose restriction to every $\mathrm{Sk}(\Delta^k_t)$, $t \in \mathcal{C}^k$, is the identity and that is continuous on Δ^k_t. Then with probability one, almost every point of p is contained in $\cup_{t \in \mathcal{C}^k} \Delta^k_t$; only the cut points are lost. To deal with this, let σ°_k denote the map acting on subsets of $\mathcal{T}$ which removes the points x such that $(t, x) \in \mathcal{P}$

and $t \in \mathcal{C}^k$; so the image of σ_k° is contained in $\cup_{t\in\mathcal{C}^k}\Delta_t^k$. Then, we set $\phi_k := \phi_k^\circ \circ \sigma_k^\circ$ and $p_k := \phi_k(p)$ is a union of disjoint open paths $\mathcal{G}_k$.

By the recursive construction in Proposition 3, understanding how the path p gets mapped into $\mathcal{G}_k$ by ϕ_k reduces to understanding one step of the transformation, that is how p gets mapped into $\mathcal{G}_1 = \mathrm{cut}(\mathcal{T}, V_1)$ by ϕ_1. The following result for $k = 1$ has been proved in [13]. It is the basis of the one-path reversal of the next section, and is used in Section 3.5 to derive the distribution of $p_k = \phi_k(p)$.

Proposition 5 (Distribution of p_1, [13])**.** *Almost surely, there exist $M_1, M_2 \geq 0$ and two finite sequences of elements of $\mathcal{C}^1$:*

$$0 < t_{1,0} < t_{1,1} < \cdots < t_{1,M_1} \qquad \text{and} \qquad 0 < t_{2,0} < t_{2,1} < \cdots < t_{2,M_2},$$

which are all distinct except that $t_{1,M_1} = t_{2,M_2}$, and there exist two sequences of points $(a_1'(m))_{0\leq m\leq M_1}$ and $(a_2'(m))_{0\leq m\leq M_2}$ satisfying $a_i'(0) = \xi_i$ and $a_i'(m) \in \Delta^1_{t_{i,m}}$ for $0 \leq m \leq M_i$ such that

$$p_1 = \phi_1(p) = \bigcup_{m=0}^{M_1-1}]\!]a_1'(m), x_1'(m)[\![\cup \bigcup_{m=0}^{M_2-1}]\!]a_2'(m), x_2'(m)[\![\cup\,]\!]a_1'(M_1), a_2'(M_2)[\![. \tag{5}$$

Here, $x_i'(m)$ denotes the root of $\Delta^1_{t_{i,m}}$, for $0 \leq m < M_i$, $i = 1, 2$. Moreover, conditionally on $\mathrm{cut}(\mathcal{T}, V_1)$, we have the following description of their distributions: let $(a_i''(m))_{m\geq 0}$, $i = 1, 2$, be two independent Markov chains with the following distribution: $a_i''(0)$ is a μ-point; given $t_i''(m)$, $a_i''(m+1)$ is distributed according to the restriction of μ to $\cup_{t>t_{i,m-1}}\Delta_t^1$ and $t_i''(m+1)$ is the element $t \in \mathcal{C}^1$ such that $a_i''(m+1) \in \Delta_t^1$. Moreover, if we let (M_1'', M_2'') be the pair of smallest integers (m_1, m_2) satisfying $t_1''(m_1) = t_2''(m_2)$, then $M_i'' < \infty$ almost surely, $i = 1, 2$, and we have

$$(a_i'(m))_{0\leq m\leq M_i} \overset{d}{=} (a_i''(m))_{0\leq m\leq M_i''}.$$

2.4 One-path reverse transformation and the 1-shuffle tree

As we have just seen, the cut tree $\mathcal{G}_1 = \mathrm{cut}(\mathcal{T}, V_1)$ has a distinguished path along which were grafted the subtrees that were pruned. Quite intuitively, the reversal should consist in taking that distinguished path, removing it, and reattaching the connected components thereby created back where they used to be in order to make a new tree. One should keep the intuition of Proposition 5 in mind; however, we do not want to assume that the reversal be performed on a tree that can indeed be obtained as $\mathrm{cut}(\mathcal{T}, V_1)$ for some $\mathcal{T}$ and V_1 and the results are stated in distribution.

Let $\mathcal{H}$ be a Brownian CRT rooted at ρ, and U_1 a point of $\mathcal{H}$ sampled according to the mass measure ν. By Proposition 1, $\mathcal{H}$ has the same distribution as $\mathrm{cut}(\mathcal{T}, V_1)$, where $\mathcal{T}$ is a Brownian CRT and V_1 is

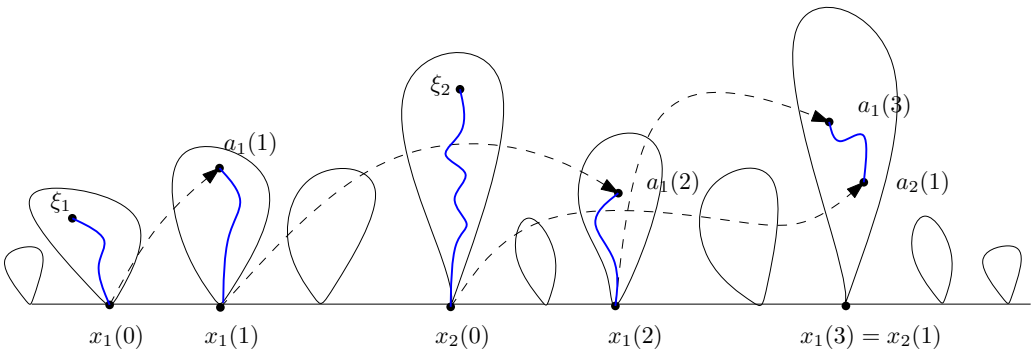

Figure 4: *An example with $\mathscr{I}(1,2) = 3$, $\mathscr{I}(2,1) = 1$ and $\mathrm{mg}(1,2) = 4$. The dashed lines indicate the point identifications for the root of the relevant subtrees.*

sampled in $\mathcal{T}$ according to the mass measure. Furthermore, we have constructed in [13] a real tree $\mathcal{Q}_1 = \mathrm{shuff}(\mathcal{H}, U_1)$ such that the following identity in distribution holds:

$$\big(\mathrm{shuff}(\mathcal{H}, U_1), \mathcal{H}\big) \overset{d}{=} \big(\mathcal{T}, \mathrm{cut}(\mathcal{T}, V_1)\big). \tag{6}$$

In particular, (6) implies that $\mathcal{Q}_1$ and $\mathcal{T}$ have the same distribution.

Let us now recall briefly the construction of $\mathcal{Q}_1$. Let $\{F_i^\circ\}_{i\geq 1}$ be the collection of connected components of $\mathcal{H} \setminus [\![\rho, U_1]\!]$ which have positive ν-mass (this is indeed countable). For each F_i°, there exists a unique point $x \in [\![\rho, U_1]\!]$ such that $F_x := F_i^\circ \cup \{x\}$ is connected in $\mathcal{H}$. Let $\mathbf{B}$ be the set of those points x. For two points x, x' of $\mathbf{B}$, we write $x \succ x'$ if $d_{\mathcal{H}}(\rho, x) > d_{\mathcal{H}}(\rho, x')$. Then for each $x \in \mathbf{B}$, we associate an attaching point A_x, which is independent and sampled according to the restriction of ν to $\cup_{x' \succ x} F_{x'}$.

By Proposition 5, in a space where we would have $\mathcal{H} = \mathrm{cut}(\mathcal{T}, V_1)$ and where $[\![\rho, U_1]\!]$ would be the distinguished path created, it would be possible to couple these choices with the cutting procedure in such a way that each A_x corresponds to the location in the initial tree $\mathcal{T}$ where F_x was detached from. Informally, if we were to glue these F_x back at A_x, we should obtain $\mathcal{T}$ back. So these choices are the correct ones, but nevertheless this transformation is *a priori* not well-defined, contrary to the discrete one (see [13, Section 3]).

The formal justification of this reverse transformation requires first to verify that the distance between two independent ν-points is a.s. well-defined (here only finitely many reattaching operations are needed), and then to construct $\mathrm{shuff}(\mathcal{H}, U_1)$ as the continuum random tree corresponding to the matrix of distances between a sequence of i.i.d. ν-points. In other words, unlike $\mathrm{cut}(\mathcal{T}, V_1)$, we do not construct $\mathrm{shuff}(\mathcal{H}, U_1)$ by actually reassembling pieces of $\mathcal{H}$, but we construct a tree that has the same metric properties.

Proposition 6 (Construction of $\mathrm{shuff}(\mathcal{H}, U_1)$, [13])**.** *Let $(\eta_i)_{i\geq 1}$ be a sequence of independent points of $\mathcal{H}$ sampled according to the mass measure ν. For each $i \geq 1$, let $(a_i(m))_{m\geq 0}$ be a sequence of points obtained as follows: $a_i(0) = \eta_i$; inductively for $m \geq 1$, let $x_i(m-1)$ be the element of $\mathbf{B}$ such that $a_i(m-1) \in F_{x_i(m-1)}$, and set $a_i(m) = A_{x_i(m)}$. Then for each pair $i \neq j$, the following quantity is almost surely finite:*

$$\mathscr{I}(i,j) := \inf\{m \geq 0 : \exists\, m' \geq 0 \text{ such that } x_i(m) = x_j(m')\};$$

and if we let

$$\begin{aligned}\gamma(i,j) := &\sum_{\ell=0}^{\mathscr{I}(i,j)-1} d_{\mathcal{H}}\big(a_i(\ell), x_i(\ell)\big) + \sum_{m=0}^{\mathscr{I}(j,i)-1} d_{\mathcal{H}}\big(a_j(m), x_j(m)\big)\\ &+ d_{\mathcal{H}}\big(a_i(\mathscr{I}(i,j)), a_j(\mathscr{I}(j,i))\big),\end{aligned} \tag{7}$$

We think of the matrix $(\gamma(i,j))_{i,j\geq 1}$ as the distance matrix between the points $(\eta_i)_{i\geq 1}$ in a new metric space. Then $(\gamma(i,j))_{i,j\geq 1}$ defines a CRT, which we root at η_1 and denote by $\mathrm{shuff}(\mathcal{H}, U_1)$. Moreover, we have

$$\big(\mathrm{shuff}(\mathcal{H}, U_1), \mathcal{H}\big) \overset{d}{=} \big(\mathcal{T}, \mathrm{cut}(\mathcal{T}, V_1)\big). \tag{8}$$

Note that only the points $(\eta_i)_{i\geq 1}$ are kept from $\mathcal{H}$. The other ones are constructed by the metric space completion. Furthermore, the very construction of the tree $\mathcal{Q}_1 = \mathrm{shuff}(\mathcal{H}, U_1)$ implies that, if one denotes by ν_1 its mass measure, then $(\eta_i)_{i\geq 1}$ is a sequence of i.i.d. ν_1-points in $\mathcal{Q}_1$. Observe also that $\gamma(i,j)$ corresponds to the distance between η_i and η_j after grafting all the $F_{x_i(\ell)}$ at $a_i(\ell+1)$ for $\ell < \mathscr{I}(i,j)$ and all the $F_{x_j(m)}$ at $a_j(m+1)$ for $m < \mathscr{I}(j,i)$ (see Figure 4). Again, one may think of a coupling where the points η_i, $i \geq 1$, would be chosen by sampling $(\xi_i)_{i\geq 1}$ in $\mathcal{T}$. Almost surely, all these points are still in $\mathrm{cut}(\mathcal{T}, V_1)$.

2.5 Multiple-paths reversal and the k-shuffle tree

Once $\mathrm{shuff}(\mathcal{H}, U_1)$ has been properly defined, the k-shuffle tree $\mathcal{Q}_k = \mathrm{shuff}(\mathcal{H}, U_1, \cdots, U_k)$ is then defined by induction. Suppose that we have constructed $\mathcal{Q}_{k-1}$ from $\mathcal{H}$ for some $k \geq 2$. Let $\tilde{T}_k^\circ$ be the component of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \cdots, U_{k-1})$ containing U_k, and let $\tilde{T}_k$ be the completion of $\tilde{T}_k^\circ$. If we write $\tilde{\mathrm{m}}_k := \mu(\tilde{T}_k)$, then $\tilde{\mathrm{m}}_k^{-1/2}\tilde{T}_k$ is distributed as a standard Brownian CRT and $\mathrm{shuff}(\tilde{T}_k, U_k)$ is thus well-defined by Proposition 6. Now let $\tilde{\mathcal{H}}_k$ be the tree obtained from $\mathcal{H}$ by replacing $\tilde{T}_k$ with $\mathrm{shuff}(\tilde{T}_k, U_k)$. Then we define $\mathcal{Q}_k := \mathrm{shuff}(\tilde{\mathcal{H}}_k, U_1, \cdots, U_{k-1})$. The following is a continuous analog of Proposition 4.8 in [13].

Proposition 7 (Distribution of the k-shuffle tree)**.** *For each $k \geq 1$, we have*

$$\big(\mathrm{shuff}(\mathcal{H}, U_1, \cdots, U_k), \mathcal{H}\big) \overset{d}{=} \big(\mathcal{T}, \mathrm{cut}(\mathcal{T}, V_1, \cdots, V_k)\big). \tag{9}$$

Proof. We proceed by induction on $k \geq 1$. The base case $k = 1$ is (8) from Proposition 6. Assume now that (9) holds for all natural numbers up to $k - 1 \geq 1$. By the scaling property, $\tilde{\mathrm{m}}_k^{-1/2}\tilde{T}_k$, equipped with the restriction of ν to $\tilde{T}_k$, is distributed as $\mathcal{H}$, and is independent of $\mathcal{H} \setminus \tilde{T}_k$. Thus, we can apply the induction hypothesis to find that

$$\big(\mathrm{shuff}(\tilde{T}_k, U_k), \tilde{T}_k\big) \overset{d}{=} \big(\Delta^k_{\tau_k}, \mathrm{cut}(\Delta^k_{\tau_k}, V_k)\big), \tag{10}$$

as $\Delta^k_{\tau_k}$ has the same distribution as $\tilde{T}_k$. In particular, all four trees in (10) are Brownian CRTs and $\mathrm{shuff}(\tilde{T}_k, U_k)$ and $\tilde{T}_k$ have the same distribution, and we deduce from the definition of $\tilde{\mathcal{H}}_k$ that

$$\big(\tilde{\mathcal{H}}_k, \mathrm{shuff}(\tilde{T}_k, U_k)\big) \overset{d}{=} \big(\mathcal{H}, \tilde{T}_k\big). \tag{11}$$

It follows that $\tilde{\mathcal{H}}_k$ and $\mathcal{H}$ have the same distribution. Then by the induction hypothesis, we have

$$\big(\mathcal{Q}_k, \tilde{\mathcal{H}}_k\big) \overset{d}{=} \big(\mathcal{T}, \mathcal{G}_{k-1}\big). \tag{12}$$

Now $\Delta^k_{\tau_k}$ is the connected component of $\mathcal{G}_{k-1} \setminus S_{k-1}$ containing the leaf labelled as k. It is thus obtained in the same way as $\tilde{T}_k$ from $\mathcal{H}$. As a consequence, $(\mathcal{H} \setminus \tilde{T}_k, \tilde{T}_k)$ and $(\mathcal{G}_{k-1} \setminus \Delta^k_{\tau_k}, \Delta^k_{\tau_k})$ have the same distribution. Combining this with (11) and (12), we obtain

$$\big(\mathcal{Q}_k, \tilde{\mathcal{H}}_k \setminus \mathrm{shuff}(\tilde{T}_k, U_k), \mathrm{shuff}(\tilde{T}_k, U_k)\big) \overset{d}{=} \big(\mathcal{T}, \mathcal{G}_{k-1} \setminus \Delta^k_{\tau_k}, \Delta^k_{\tau_k}\big).$$

The transformation from $\tilde{T}_k$ to $\mathrm{shuff}(\tilde{T}_k, U_k)$ only involves sampling random points in $\tilde{T}_k$, and is therefore independent of the transformation from $\tilde{\mathcal{H}}_k$ to $\mathcal{Q}_k$. Similarly the transformations from $\mathcal{T}$ to $\mathcal{G}_{k-1}$ and the one from $\Delta^k_{\tau_k}$ to $\mathrm{cut}(\Delta^k_{\tau_k}, V_k)$ are also independent. Then it follows from (10) that

$$\big(\mathcal{Q}_k, \tilde{\mathcal{H}}_k \setminus \mathrm{shuff}(\tilde{T}_k, U_k), \mathrm{shuff}(\tilde{T}_k, U_k), \tilde{T}_k\big) \overset{d}{=} \big(\mathcal{T}, \mathcal{G}_{k-1} \setminus \Delta^k_{\tau_k}, \Delta^k_{\tau_k}, \mathrm{cut}(\Delta^k_{\tau_k}, V_k)\big)$$

Finally, observe that $\tilde{\mathcal{H}}_k \setminus \mathrm{shuff}(\tilde{T}_k, U_k) = \mathcal{H} \setminus \tilde{T}_k$ and that $\mathcal{G}_{k-1} \setminus \Delta^k_{\tau_k} = \mathcal{G}_k \setminus \mathrm{cut}(\Delta^k_{\tau_k}, V_k)$. This yields

$$\big(\mathcal{Q}_k, \mathcal{H} \setminus \tilde{T}_k, \tilde{T}_k\big) \overset{d}{=} \big(\mathcal{T}, \mathcal{G}_k \setminus \mathrm{cut}(\Delta^k_{\tau_k}, V_k), \mathrm{cut}(\Delta^k_{\tau_k}, V_k)\big),$$

which entails that (9) holds for k, and completes the proof. □

3 Convergence of k-shuffle trees and the shuffle tree

3.1 The shuffle tree

In this section, we prove the following result, which constitutes the foundations of the formal definition of the shuffle tree.

Theorem 8 (Convergence of the k-shuffle trees)**.** *For a.e. Brownian CRT $\mathcal{H}$, the limit of the sequence $(\mathrm{shuff}(\mathcal{H}, U_1, \cdots, U_k))_{k\geq 1}$ exists almost surely in the Gromov–Prokhorov topology.*

The sequence of leaves $(U_i)_{i\geq 1}$ that is used influences the limit: in particular, it determines which subtrees are fringes and in which direction they are sent to. Still, in the same way that $\mathrm{cut}(\mathcal{T})$ does depend on the cutting procedure, we denote the limit by $\mathrm{shuff}(\mathcal{T})$, although it does depend on the sequence $(U_i)_{i\geq 1}$. However, $(U_i)_{i\geq 1}$ *does not* contain all the randomness used in the construction. In Section 4, we construct a specific sequence of leaves which emphasizes the randomness hidden in the construction, justifying the claim in the introduction that the sequence $(A_x, x \in \mathrm{Br}(\mathcal{H}))$ is all that one needs.

The following is a direct consequence of Theorem 8, and justifies the claim that $\mathrm{shuff}(\,\cdot\,)$ is indeed the reverse transformation of $\mathrm{cut}(\,\cdot\,)$.

Corollary 9. *We have the following identity in distribution:*

$$\big(\mathrm{shuff}(\mathcal{H}), \mathcal{H}\big) \stackrel{d}{=} \big(\mathcal{T}, \mathrm{cut}(\mathcal{T})\big). \tag{13}$$

Proof. Let f and g be two bounded real-valued functions that are continuous in the Gromov–Prokhorov topology. Recall the notation $\mathcal{G}_k = \mathrm{cut}(\mathcal{T}, V_1, \cdots, V_k)$. By (9), we have

$$\mathbb{E}[f(\mathcal{Q}_k) \cdot g(\mathcal{H})] = \mathbb{E}[f(\mathcal{T}) \cdot g(\mathcal{G}_k)].$$

By Proposition 4 and the dominated convergence theorem, the right-hand side above converges to $\mathbb{E}[f(\mathcal{T}) \cdot g(\mathrm{cut}(\mathcal{T}))]$, as $k \to \infty$. Similarly, by Theorem 8, the left-hand side converges to $\mathbb{E}[f(\mathrm{shuff}(\mathcal{H})) \cdot g(\mathcal{H})]$. Therefore,

$$\mathbb{E}[f(\mathrm{shuff}(\mathcal{H}) \cdot g(\mathcal{H})] = \mathbb{E}[f(\mathcal{T}) \cdot g(\mathrm{cut}(\mathcal{T}))].$$

Since f and g were arbitrary, this entails (13). □

Let $(\eta_i)_{i\geq 1}$ be the sequence of independent points of $\mathcal{H}$ in Proposition 6, which is independent of the sequence $(U_i)_{i\geq 1}$. Recall the random variable $\gamma(i, j)$, which is the distance between η_i and η_j in $\mathcal{Q}_1$. Part of the statement of Proposition 6 says that $(\eta_i)_{i\geq 1}$ is a a family of i.i.d. uniform points in $\mathcal{Q}_1$. Because of the inductive definition of $\mathcal{Q}_k$, the sequence $(\eta_i)_{i\geq 1}$ remains an i.i.d. uniform family in each $\mathcal{Q}_k$. Let us denote by $\gamma_k(i, j)$ the distance between η_i and η_j in $\mathcal{Q}_k$ for $k \geq 1$. The main tool towards Theorem 8 is the following proposition:

Proposition 10. *For each $i, j \geq 1$, $\gamma_k(i, j) \to \gamma_\infty(i, j) < \infty$ almost surely as $k \to \infty$.*

Let us first explain why this entails Theorem 8.

Proof of Theorem 8. Observe that by Proposition 7, for each $k \geq 1$, $\mathcal{Q}_k$ is distributed as the Brownian CRT $\mathcal{H}$. Thus, $(\gamma_k(i, j))_{i,j\geq 1}$ and $(d_{\mathcal{H}}(\eta_i, \eta_j))_{i,j\geq 1}$ have the same distribution for each $k \geq 1$. It follows from Proposition 10 that the limit matrix $(\gamma_\infty(i, j))_{i,j\geq 1}$ is also distributed as $(d_{\mathcal{H}}(\eta_i, \eta_j))_{i,j\geq 1}$. In other words, $(\gamma_\infty(i, j))_{i,j\geq 1}$ has the distribution of the distance matrix of the Brownian CRT. In particular, for each $n \geq 1$, $(\gamma_\infty(i, j))_{1\leq i,j\leq n}$ defines an n-leaf real tree $\mathcal{R}_n$ and the family $(\mathcal{R}_n)_{n\geq 1}$ is *consistent* and *leaf-tight* (see [3]), which means that $(\mathcal{R}_n)_{n\geq 1}$ admits a representation as a continuum random tree.

Observe that all these are still true a.s. conditionally on $\mathcal{H}$. More precisely, by Proposition 10, for $\mathbb{P}$-almost every $\mathcal{H}$, as $k \to \infty$,

$$\big(\gamma_k(i,j), i,j \le n\big) \xrightarrow{a.s.} \big(\gamma_\infty(i,j), i,j \le n\big). \tag{14}$$

for each $n \ge 1$, conditionally on $\mathcal{H}$. For those $\mathcal{H}$ for which (14) holds, the family $(\mathcal{R}_n)_{n\ge1}$ is a.s. consistent and leaf-tight, conditionally on $\mathcal{H}$. Let $\mathcal{R}_\infty(\mathcal{H})$ be the CRT representation of this family. Then by definition, (14) entails the Gromov–Prokhorov convergence of $\mathrm{shuff}(\mathcal{H}, U_1, \cdots, U_k)$ to $\mathcal{R}_\infty(\mathcal{H})$. □

The remainder of the section is devoted to proving Proposition 10. Since $(\eta_i)_{i\ge1}$ is an i.i.d. sequence, it suffices to consider the case $i = 1, j = 2$.

3.2 A series representation for $\gamma_k(1,2)$

The idea behind the formal definition of Proposition 6 is to leverage Proposition 5 as follows: if $\mathcal{H}$ were $\mathrm{cut}(\mathcal{T}, V_1)$, for some $\mathcal{T}$ and V_1, and the distinguished path were the one between the root and $U_1 \in \mathcal{H}$, then the image of a path between two points ξ_1 and ξ_2 in $\mathcal{T}$ would now go through a number of subtrees of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}, U_1)$, and in every such tree it would go between two points which are uniform. We now go further and give such a representation for $\gamma_k(1,2)$ as a sum where we specify the distributions of the trees and points involved.

The masses of these trees are of prime importance, and we let

$$\mathcal{S}^\downarrow = \left\{(x_0, x_1, \cdots) : x_0 \ge x_1 \ge \cdots \ge 0; \sum_{i\ge0} x_i \le 1\right\}$$

be the space of mass partitions, equipped with the usual ℓ_1-norm $\|\cdot\|_1$. If $\mathbf{x} = (x_1, x_2, \cdots) \in \mathcal{S}^\downarrow$, then the *length* of $\mathbf{x}$ is defined to be the smallest index n such that $x_n = 0$, which may well be infinite. And we denote by $\mathcal{S}^\downarrow_f$ the subset of $\mathcal{S}^\downarrow$ which consists of the elements of finite length.

Recall the definition of $\gamma_1(1,2)$ in (7). The trees involved there are the components $F_{x_i(n)}$ rooted at $x_i(n)$, for $n \ge 0$ and $i = 1, 2$. Let ϖ denote the distribution of the rearrangement of

$$\{\nu(F_{x_1(n)}), 0 \le n \le \mathscr{I}(1,2)\} \cup \{\nu(F_{x_2(n)}), 0 \le n \le \mathscr{I}(2,1) - 1\}$$

in decreasing order. Then ϖ is a probability measure supported on $\mathcal{S}^\downarrow_f$.

Lemma 11 (Representation of $\gamma_k(1,2)$)**.** *For each $k \ge 1$, there exists a finite sub-collection of the masses of the components of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \cdots, U_k)$ denoted by $\mathbf{m}_k = (m_{k,n})_{0\le n\le N_k} \in \mathcal{S}^\downarrow_f$, and a sequence of positive real numbers $(\mathrm{R}^k_n, 0 \le n \le N_k)$ such that*

$$\gamma_k(1,2) = \sum_{n=0}^{N_k} \sqrt{m_{k,n}}\, \mathrm{R}^k_n, \tag{15}$$

where $(\mathrm{R}^k_n)_{n\ge0}$ is a sequence of i.i.d. copies of a Rayleigh random variable that is independent of $\mathbf{m}_k$ and N_k. Moreover, $(\mathbf{m}_k)_{k\ge1}$ is a Markov chain with initial law ϖ and the following transitions:

- *with probability $1 - \|\mathbf{m}_k\|_1$, $\mathbf{m}_{k+1} = \mathbf{m}_k$, and*
- *for $0 \le n \le N_k$, with probability $m_{k,n}$, $\mathbf{m}_{k+1}$ is obtained by replacing in $\mathbf{m}_k$ the element $m_{k,n}$ by $m_{k,n} \cdot \widetilde{\mathbf{m}}$, where $\widetilde{\mathbf{m}}$ has distribution ϖ, and then resorting the sequence in decreasing order.*

Proof. Let us first consider the case $k = 1$. Let $\{\mathrm{T}_{1,n}, 0 \le n \le N_1\}$ be the collection $\{F_{x_1(m)}, 0 \le m \le \mathscr{I}(1,2)\} \cup \{F_{x_2(m)}, 0 \le m \le \mathscr{I}(2,1)\}$ sorted in such a way that $\mathbf{m}_1 := (\nu(\mathrm{T}_{1,n}), 0 \le n \le N_1)$ is nonincreasing. We denote by $\mathbf{F}_1 = (\mathrm{T}_{1,n}, n \ge 0)$ and $\mathbf{m}_1 = (m_{1,n}, n \ge 0)$. Then the distribution of $\mathbf{m}_1$ is ϖ. For each $0 \le n \le N_1$, define (see Figure 4)

$$D_n^1 = \begin{cases} d_{\mathcal{H}}(x_1(m), a_1(m)), & \text{if } \mathrm{T}_{1,n} = F_{x_1(m)} \text{ for } m = 0, 1, \cdots, \mathscr{I}(1,2) - 1; \\ d_{\mathcal{H}}(x_2(m), a_2(m)), & \text{if } \mathrm{T}_{1,n} = F_{x_2(m)} \text{ for } m = 0, 1, \cdots, \mathscr{I}(2,1) - 1; \\ d_{\mathcal{H}}(a_1(\mathscr{I}(1,2)), a_2(\mathscr{I}(2,1))), & \text{otherwise.} \end{cases}$$

Then, $m_{1,n} > 0$ for $0 \le n \le N_1$, and we set $\mathrm{R}_n^1 := D_n^1 m_{1,n}^{-1/2}$. By (7), this definition immediately yields

$$\gamma_1(1,2) = \sum_{n=0}^{N_1} \sqrt{m_{1,n}}\, \mathrm{R}_n^1. \tag{16}$$

This is (15) for $k = 1$. For the distribution of $(\mathrm{R}_n^1)_{n \ge 1}$, we need the following fact, whose proof is given in Appendix A.

Lemma 12 (Scaling property)**.** *For $0 \le n \le N_1$, let T_n^* be the rescaled metric space $m_{1,n}^{-1/2}\mathrm{T}_{1,n}$, equipped with the (probability rescaled) restriction of ν to $\mathrm{T}_{1,n}$. Then for each $j \ge 1$, on the event $\{N_1 = j\}$, $(\mathrm{T}_n^*)_{0 \le n \le N_1}$ is a sequence of j independent copies of a Brownian CRT.*

Let us recall that by definition each $a_i(m)$ is distributed as ν restricted to $F_{x_i(m)}$. We also recall that if η, η' are two independent points of $\mathcal{H}$ sampled according to ν, then $d_{\mathcal{H}}(\eta, \eta')$ (resp. the distance of η from the root) is Rayleigh distributed. Then it follows from Lemma 12 that $(\mathrm{R}_n^1)_{0 \le n \le N_1}$ is a sequence of i.i.d. Rayleigh random variables. This proves the statement for $k = 1$, which is our base case.

Now we proceed to prove the induction step, and assume that we almost surely have the desired representation for all natural numbers up to $k \ge 1$. In particular, there exists a sequence $\mathbf{F}_k = (\mathrm{T}_{k,n})_{0 \le n \le N_k}$ which is a finite sub-collection of the connected components of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \cdots, U_k)$ such that $\mathbf{m}_k := (\nu(\mathrm{T}_{k,n}))_{0 \le n \le N_k}$ is non-increasing. Moreover, we suppose that (15) holds for k, where for each $0 \le n \le N_k$, $\sqrt{m_{k,n}}\mathrm{R}_n^k$ is either the distance between two independent ν-points of $\mathrm{T}_{k,n}$ or the distance between a ν-point and the root of $\mathrm{T}_{k,n}$.

Recall that $\mathcal{Q}_{k+1}$ is defined to be $\mathrm{shuff}(\tilde{\mathcal{H}}_{k+1}, U_1, \cdots, U_k)$, where $\tilde{\mathcal{H}}_{k+1}$ is obtained from $\mathcal{H}$ by replacing the connected component $\tilde{T}_{k+1}$ of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \ldots, U_k)$ which contains U_{k+1}, by $\tilde{S} := \mathrm{shuff}(\tilde{T}_{k+1}, U_{k+1})$. The real tree $\tilde{\mathcal{H}}_{k+1}$ is a Brownian CRT with mass measure $\tilde{\nu}_{k+1}$, and by the induction hypothesis, there exists a sequence $\hat{\mathbf{F}}_k = (\hat{\mathrm{T}}_{k,n}, 0 \le n \le \hat{N}_k)$ which consists in a finite sub-collection of the components of $\tilde{\mathcal{H}}_{k+1} \setminus \mathrm{Span}(\tilde{\mathcal{H}}_{k+1}; U_1, \cdots, U_k)$ rearranged in decreasing order of their masses, such that

$$\gamma_{k+1}(1,2) = \sum_{n=0}^{\hat{N}_k} \sqrt{\hat{m}_{k,n}}\, \hat{\mathrm{R}}_n^k, \tag{17}$$

where for each $0 \le n \le \hat{N}_k$, $\hat{m}_{k,n} = \mu(\hat{\mathrm{T}}_{k,n})$ and $\sqrt{\hat{m}_{k,n}}\hat{\mathrm{R}}_n^k$ is either the distance between two uniform independent points of $\hat{\mathrm{T}}_{k,n}$ or the distance between a uniform point and the root of $\hat{\mathrm{T}}_{k,n}$. However, we are after a representation in terms of the connected components of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \ldots, U_{k+1})$.

Note that $\tilde{S}$ is the only component of $\tilde{\mathcal{H}}_{k+1} \setminus \mathrm{Span}(\tilde{\mathcal{H}}_{k+1}; U_1, \cdots, U_k)$ that is not a component of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \ldots, U_{k+1})$. So if $\tilde{S}$ does not appear in $\hat{\mathbf{F}}_k$, then by construction we have $\hat{\mathbf{F}}_k = \mathbf{F}_k$ where $\mathbf{F}_k = (\mathrm{T}_{k,n})_{0 \le n \le N_k}$ is a finite sub-collection of the connected components of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \cdots, U_k)$ such that $\mathbf{m}_k := (\nu(\mathrm{T}_{k,n}))_{0 \le n \le N_k}$ is non-increasing for which (15) holds, with the additional distributional properties we are after. Furthermore, in that case, we have $\gamma_{k+1}(1,2) = \gamma_k(1,2)$. It thus suffices to take $\mathbf{F}_{k+1} = \mathbf{F}_k$ and $\mathrm{R}_n^{k+1} = \mathrm{R}_n^k$ for each n. This case occurs precisely if U_{k+1} does not fall in any of the subtrees of $\mathbf{F}_k$, which happens with probability $1 - \|\mathbf{m}_k\|_1$, since U_{k+1} is ν-distributed.

If, on the other hand, $\tilde{S} = \hat{\mathrm{T}}_{k,n_0}$ for some $0 \le n_0 \le \hat{N}_k$, then the representation in (17) needs to be modified. Still, since $\hat{m}_{k,n_0} = \tilde{\nu}_{k+1}(\tilde{S}) = \nu(\tilde{T}_{k+1}) = m_{k,n_0}$, the masses are correct and $\hat{\mathbf{m}}_k = \mathbf{m}_k$. So, in particular, this occurs with probability m_{k,n_0} Note also that, by definition of $\tilde{\mathcal{H}}_{k+1}$,

$$\gamma_{k+1}(1,2) - \gamma_k(1,2) = \hat{\mathrm{R}}^k_{n_0} - \mathrm{R}^k_{n_0}, \tag{18}$$

where here, $(m_{k,n_0})^{1/2}\hat{\mathrm{R}}^k_{n_0}$ is the distance in $\tilde{S}$ between either two independent $\tilde{\nu}_{k+1}$-points or between a $\tilde{\nu}_{k+1}$-point and the root. Recall that $\tilde{S} = \mathrm{shuff}(\tilde{T}_{k+1}, U_{k+1})$ is rooted at a $\tilde{\nu}_{k+1}$-point. Note also that by the scaling property, $(m_{k,n_0})^{-1/2}\tilde{T}_{k+1}$ is a Brownian CRT (and is thus distributed as $\mathcal{H}$). Therefore, we may use the induction hypothesis with $k = 1$ to obtain that there exists a sequence $\check{\mathbf{F}} = (\check{\mathrm{T}}_n, 0 \le n \le \check{N})$ consisting in a sub-collection of the connected components of $\tilde{T}_{k+1} \setminus \mathrm{Span}(\tilde{T}_{k+1}; U_{k+1})$ rearranged in the decreasing order of their masses such that

$$\hat{\mathrm{R}}^k_{n_0} = \sum_{n=0}^{\check{N}} \sqrt{\nu(\check{\mathrm{T}}_n)}\, \check{\mathrm{R}}_n, \tag{19}$$

where for each $0 \le n \le \check{N}$, $\nu(\check{T}_n)^{1/2}\check{\mathrm{R}}_n$ is either the distance in $\check{\mathrm{T}}_n$ between either two independent ν-points or between a ν-point and the root. So in particular, $(\check{\mathrm{R}}_n)_{n\ge 0}$ forms a sequence of i.i.d. copies of a Rayleigh random variable. Furthermore, $(\nu(\check{\mathrm{T}}_n)/m_{k,n_0}, 0 \le n \le \check{N})$ is an independent copy of $\mathbf{m}_1$. Then we set $\mathbf{F}_{k+1} = (\mathrm{T}_{k+1,n}, 0 \le n \le N_{k+1})$ to be the rearrangement of the collection

$$\{\hat{\mathrm{T}}_{k,n} : 0 \le n \le \hat{N}_k, n \ne n_0\} \cup \{\check{\mathrm{T}}_n : 0 \le n \le \check{N}\}$$

such that $\mathbf{m}_{k+1} := (\nu(\mathrm{T}_{k+1,n}))_{0\le n\le N_{k+1}}$ is non-increasing. Note that $\mathbf{F}_{k+1}$ is a finite sub-collection of components of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1, \cdots, U_{k+1})$. Finally, inserting (19) into (17) and then comparing with (18), we obtain (15) for $k+1$, which completes the proof. □

Proving Proposition 10 now reduces to showing that the series representation in (15) converges almost surely. First observe that conditionally on $\mathbf{m}_k$, $\gamma_k(1,2)$ is a sum of independent random variables which have Gaussian tails (see later for details). It then easily follows from classical results on concentration of measure that $\gamma_k(1,2)$ is concentrated about the conditional mean $\mathbb{E}[\gamma_k(1,2) \mid \mathbf{m}_k]$. Furthermore the width of the concentration window is controlled by the variance, which is here $O(\|\mathbf{m}_k\|_1)$. The following lemmas control the distance between $\gamma_k(1,2)$ and $\mathbb{E}[\gamma_k(1,2) \mid \mathbf{m}_k]$.

Lemma 13. *There exists some $\alpha > 0$ such that*

$$\lim_{k\to\infty} k^\alpha \|\mathbf{m}_k\|_1 = 0, \qquad \textit{almost surely.}$$

Lemma 14. *We have $\gamma_k(1,2) - \mathbb{E}[\gamma_k(1,2) \mid \mathbf{m}_k] \to 0$ almost surely as $k \to \infty$.*

The proofs of these lemmas rely on standard facts about fragmentations chains and concentration inequalities and are presented in Sections 3.3 and 3.4. From there, the last step consists in proving that $\mathbb{E}[\gamma_k(1,2) \mid \mathbf{m}_k]$ also converges almost surely.

Lemma 15. *A.s., $\mathbb{E}[\gamma_k(1,2) \mid \mathbf{m}_k]$ converges to some random variable $\gamma_\infty(1,2) < \infty$ as $k \to \infty$.*

Our approach to Lemma 15 relies on a coupling between the cutting and shuffling procedure and is given in Section 3.5.

3.3 Proof of Lemma 13: polynomial decay of the self-similar fragmentation chain

The dynamics of $(\mathbf{m}_k)_{k\geq 0}$ are quite similar to that of a self-similar fragmentation chain, and the proof of the lemma relies on classical results on the asymptotic behaviour of fragmentation processes. If we were to count only the number of *actual* jumps of the process (or equivalently the number of i's such that U_i does affect the collection of masses) then one would exactly have the state of a fragmentation chain taken at the jump times. The fact that even the i's that do not affect the chain are counted only induces a time-change that is easily controlled.

Recall that ϖ denotes the law of $\mathbf{m}_1$. Let $X(t) = (X_i(t))_{i\geq 1}$ be a self-similar fragmentation chain with index of self-similarity 1 and dislocation measure ϖ starting from the initial state $X(0) = (1, 0, 0, \cdots)$, as introduced in [6, Chapter 1]. Then, $X(t)$ jumps at rate $\|X(t)\|_1 = \sum_{i\geq 1} X_i(t)$. This chain is non-conservative since $\mathbb{P}(0 < \|\mathbf{m}_1\|_1 < 1) = 1$ for $\mathbf{m}_1$ is an a.s. finite and non-empty collection of the masses of the components of $\mathcal{H} \setminus \mathrm{Span}(\mathcal{H}; U_1)$. To compensate for the loss of mass (the i's that do not modify $(\mathbf{m}_k)_{k\geq 0}$), consider another Poisson point process Γ on $[0, \infty)$ with rate $1 - \|X(t)\|_1$ at time $t \geq 0$, and let $\theta(t)$ denote the number of jumps before time t in X and Γ combined. Then $\theta(t)$ is the number of jumps before time t of a Poisson process with rate one, and if we set $\theta^{-1}(k) := \inf\{t \geq 0 : \theta(t) \geq k\}$ be the time of the k-th point, then we have

$$(X(\theta^{-1}(k)))_{k\geq 1} \stackrel{d}{=} (\mathbf{m}_k)_{k\geq 1}. \tag{20}$$

From now on, we work on a space on which these are coupled to be equal with probability one.

Let $p^* \in (0, 1)$ be the critical exponent such that $\mathbb{E}[\sum_{i\geq 1} m_{1,i}^{p^*}] = 1$. Then

$$\mathcal{M}(t) := \sum_{i\geq 1} X_i^{p^*}(t)$$

is a positive martingale which is uniformly integrable. Denote by $\mathcal{M}(\infty)$ its a.s. limit. By Theorem 1 of [9], for every $k \geq 1$, there exists a constant $C_k \in (0, \infty)$ such that

$$\sup_{t\geq 0} t^k \cdot \mathbb{E}\left[\sum_{i\geq 1} X_i^{p^*+k}(t)\right] \leq C_k,$$

from which it follows immediately that $\sup_{t\geq 0} t^k \cdot \mathbb{E}X_1^{p^*+k}(t) \leq C_k$. Now, for any $\delta \in (0, 1)$ and $\epsilon > 0$, by Markov's inequality, we obtain at time $n \geq 0$

$$\mathbb{P}\big(X_1(n) \geq \epsilon n^{-\delta}\big) \leq \frac{n^k \cdot \mathbb{E}X_1^{p^*+k}(n)}{\epsilon^{p^*+k} n^{(1-\delta)k-\delta p^*}} \leq \frac{C_k}{\epsilon^{p^*+k} n^{(1-\delta)k-\delta p^*}}.$$

By choosing k large enough that $k > (1+\delta p^*)/(1-\delta)$, this implies that $\sum_{n\geq 1} \mathbb{P}(X_1(n) \geq \epsilon n^{-\delta}) < \infty$ and $\limsup_{n\to\infty} n^\delta X_1(n) \leq \epsilon$ almost surely, by the Borel–Cantelli lemma. Letting $\epsilon \to 0$ along a sequence, we then obtain that $n^\delta X_1(n) \to 0$ a.s., and $t^\delta X_1(t) \to 0$ a.s. as well by monotonicity. Now notice that for any $t \geq 0$,

$$\sum_{i\geq 1} X_i(t) = \sum_{i\geq 1} X_i^{1-p^*}(t) \cdot X_i^{p^*}(t) \leq X_1^{1-p^*}(t) \cdot \mathcal{M}(t).$$

Then, for any $\delta \in (0, 1)$,

$$\limsup_{t\to\infty} t^{\delta(1-p^*)} \sum_{i\geq 1} X_i(t) \leq \mathcal{M}(\infty) \cdot \limsup_{t\to\infty} (t^\delta X_1(t))^{1-p^*} = 0 \quad \text{almost surely,} \tag{21}$$

since $p^* \in (0, 1)$ as the chain is non-conservative.

Finally, we go back to $(\mathbf{m}_k)_{k\geq 1}$ using (20). By the strong law of large numbers, we have $\theta(t)/t \to 1$ almost surely, and we easily deduce that $k/\theta^{-1}(k) \to 1$ almost surely as $k \to \infty$. Using this fact together with (20) and (21), we obtain that for any $\delta \in (0,1)$,

$$\limsup_{k\to\infty} k^{\delta(1-p^*)} \|\mathbf{m}_k\|_1 = \limsup_{k\to\infty} k^{\delta(1-p^*)} \sum_{i\geq 1} X_i(\theta^{-1}(k)) = 0 \quad \text{almost surely,}$$

which completes the proof of Lemma 13.

3.4 Proof of Lemma 14: concentration of the Rayleigh variable

Let R be a random variable of Rayleigh distribution, with density $xe^{-x^2/2}$ on $\mathbb{R}_+$. Then, one easily verifies that $\mathrm{R} - \mathbb{E}[\mathrm{R}]$ is *sub-Gaussian* in the sense that there exists a constant v such that for every $\lambda \in \mathbb{R}$, one has

$$\log \mathbb{E}[e^{\lambda(\mathrm{R}-\mathbb{E}[\mathrm{R}])}] \leq \frac{\lambda^2 v}{2}.$$

(See [12, Theorem 2.1, p. 25].) We may thus apply concentration results for sub-Gaussian random variables such as those presented in Section 2.3 of [12].

For each $k \geq 1$, we have

$$\sigma_k := \gamma_k(1,2) - \widetilde{\gamma}_k = \sum_{i=1}^{N_k} \sqrt{m_{k,n}} \left(\mathrm{R}_n^k - \mathbb{E}[\mathrm{R}_n^k]\right),$$

by (15), where according to Lemma 11, $(\mathrm{R}_n^k, 1 \leq n \leq N_k)$ is a sequence of i.i.d. copies of a Rayleigh random variable. Therefore

$$\mathbb{P}\left(|\sigma_k| \geq \epsilon \,|\, \mathbf{m}_k\right) \leq 2\exp\Big(-\frac{\epsilon^2}{2v\|\mathbf{m}_k\|_1}\Big). \tag{22}$$

If $(A_k, B_k, C_k)_{k\geq 1}$ are sequence of events satisfying that $A_k \subset B_k \cup C_k$ for each $k \geq 1$, then it is elementary that $\mathbb{P}(\limsup_k A_k) \leq \mathbb{P}(\limsup_k B_k) + \mathbb{P}(\limsup_k C_k)$. Here, we take

$$A_k = \{|\sigma_k| \geq \epsilon\}, \quad B_k = A_k \cap \{\|\mathbf{m}_k\|_1 \leq k^{-\alpha}\}, \quad C_k = \{\|\mathbf{m}_k\|_1 > k^{-\alpha}\}$$

with the same α as in Lemma 13. Then $\mathbb{P}(\limsup_k C_k) = 0$ by Lemma 13. On the other hand, we deduce from (22) that

$$\sum_{k\geq 1} \mathbb{P}(B_k) = \sum_{k\geq 1} \mathbb{E}\left[\mathbb{P}\left(|\sigma_k| \geq \epsilon \,|\, \mathbf{m}_k\right) \cdot \mathbf{1}_{\{\|\mathbf{m}_k\|_1 \leq k^{-\alpha}\}}\right] \leq \sum_{k\geq 1} 2e^{-\epsilon^2 k^\alpha/(2v)} < \infty,$$

which entails that $\mathbb{P}(\limsup_k B_k) = 0$ by the Borel–Cantelli lemma. Hence, $\mathbb{P}\left(\limsup_k A_k\right) = 0$, which means $\limsup_k |\sigma_k| < \epsilon$ almost surely. Since $\epsilon > 0$ was arbitrary, the proof of Lemma 14 is now complete.

3.5 Proof of Lemma 15: a coupling via cut trees

Let us recall the notations before Proposition 5. There are two μ-points ξ_1, ξ_2 in the Brownian CRT $\mathcal{T}$, and $p :=]\!]\xi_1, \xi_2[\![$ is the path in $\mathcal{T}$ between these two points. We denote by $D := d_{\mathcal{T}}(\xi_1, \xi_2)$ the length of this path. Let $\mathcal{G}_k = \mathrm{cut}(\mathcal{T}, V_1, \cdots, V_k)$ be the k-cut tree. Recall that, up to the finitely many cut points that are lost, p_k is the image of p by the canonical embedding ϕ_k from $\cup_{t\in\mathcal{C}^k} \Delta_t^k$ into $\mathcal{G}_k$. We have the following representation of the distance D, which is an analog of Lemma 11.

Lemma 16. *For each $k \geq 1$, there exists some $\mathbf{m}'_k = (m'_{k,n})_{0\leq n\leq N'_k} \in \mathcal{S}^{\downarrow}_f$, which is a sub-collection of the masses of $\{\Delta^k_t, t \in \mathcal{C}^k\}$ such that*

$$D = \sum_{n=0}^{N'_k} \sqrt{m'_{k,n}}\, \mathrm{B}^k_n, \tag{23}$$

where $(\mathrm{B}^k_n)_{n\geq 0}$ is an i.i.d. sequence of Rayleigh random variables, independent of $\mathbf{m}'_k$ and N'_k. Moreover, $(\mathbf{m}'_k)_{k\geq 1}$ has the same distribution as $(\mathbf{m}_k)_{k\geq 1}$.

Proof. For each $k \geq 1$, the injection ϕ_k is an isometry on each $\mathrm{Sk}(\Delta^k_t)$. Thus, we have $D = \ell(p_k)$ for each $k \geq 1$. Let us show that $\ell(p_k)$ can be written as the right-hand side in (23).

We proceed by induction on $k \geq 1$. The base case $k = 1$ is a consequence of Proposition 5. Let $\mathbf{F}'_1 = (\mathrm{T}'_{1,n}, 0 \leq n \leq N'_1)$ be the vector consisting of the elements of the collection

$$\{\Delta^1_{t_{1,m}}, 0 \leq m \leq M_1\} \cup \{\Delta^1_{t_{2,m}}, 0 \leq m \leq M_2 - 1\},$$

arranged in the decreasing order of their masses $m'_{1,n} := \mu(\mathrm{T}'_{1,n})$. By comparing Propositions 5 and 6, we see that $\mathbf{F}'_1$ has the same distribution as $\mathbf{F}_1$, since $\mathcal{G}_1$ is a Brownian CRT (Proposition 1) and given $\mathcal{G}_1$ the sequences $(a'_i(m))_{m\geq 0}$, $i \geq 1$, are sampled in the same way as $(a_i(m))_{m\geq 0}$ given $\mathcal{H}$. As a consequence, $\mathbf{m}'_1 := (m'_{1,n}, 0 \leq n \leq N'_1)$ has the same distribution as $\mathbf{m}_1$. Moreover, by Lemma 12, each $(m'_{1,n})^{-1/2}\mathrm{T}'_{1,n}$ is an independent copy of a Brownian CRT. Define

$$D^1_n := \begin{cases} d_{\mathcal{T}}(x'_1(m), a'_1(m)) & \text{if } \mathrm{T}'_{1,n} = \Delta^1_{t_{1,m}} \text{ for } m = 0, 1, \cdots, M_1 - 1; \\ d_{\mathcal{T}}(x'_2(m), a'_2(m)), & \text{if } \mathrm{T}'_{1,n} = \Delta^1_{t_{2,m}} \text{ for } m = 0, 1, \cdots, M_2 - 1; \\ d_{\mathcal{T}}(a'_1(M_1), a'_2(M_2)), & \text{otherwise,} \end{cases}$$

and set $\mathrm{B}^1_n := D^1_n/(m'_{1,n})^{1/2}$. Then B^1_n, $0 \leq n \leq N'_1$, are independent from each other. Also, it follows from (5) that

$$\ell(p_1) = \sum_{n=0}^{N'_1} (m'_{1,n})^{1/2}\, \mathrm{B}^1_n.$$

Suppose now that for all natural numbers up to some $k \geq 1$ there exist some $\mathbf{F}'_k = (\mathrm{T}'_{k,n}, 0 \leq k \leq N'_k)$, the elements of which form a sub-collection of $(\Delta^k_t, t \in \mathcal{C}^k)$, such that $\mathbf{m}'_k = (m'_{k,n})_{0\leq n\leq N'_k} := (\mu(\mathrm{T}'_{k,n}))_{0\leq n\leq N'_k}$ is non-increasing and that

$$\ell(p_k) = \sum_{n=0}^{N'_k} \sqrt{m'_{k,n}}\, \mathrm{B}^k_n, \tag{24}$$

where $(m_{k,n})^{1/2}\mathrm{B}^k_n$ is the distance between two points, say u^k_n and v^k_n, in $\mathrm{T}'_{k,n}$, and u^k_n is a μ-point in $\mathrm{T}'_{k,n}$ while v^k_n is either the root of $\mathrm{T}'_{k,n}$ or another μ-point independent of u^k_n. Recall that by Proposition 3, $\mathcal{G}_{k+1}$ can be obtained from $\mathcal{G}_k$ by replacing $\Delta^k_{\tau_{k+1}}$ with $\mathrm{cut}(\Delta^k_{\tau_{k+1}}, V_{k+1})$.

Suppose first that $\Delta^k_{\tau_{k+1}}$ does not appear in $\mathbf{F}'_k$, which happens with probability $1 - \|\mathbf{m}'_k\|_1$. Then, all the components of $\mathbf{F}'_k$ are actually elements of $\{\Delta^{k+1}_t, t \in \mathcal{C}^{k+1}\}$ (see Figure 3) and it suffices to take $\mathbf{F}'_{k+1} = \mathbf{F}'_k$ and $\mathrm{B}^{k+1}_n = \mathrm{B}^k_n$ for each n. So in this case, the representation in (23) for $k + 1$ follows trivially from (24).

Suppose now that $\Delta^k_{\tau_{k+1}} = \mathrm{T}'_{k,n_0}$ for some $0 \leq n_0 \leq N'_k$, which occurs with probability m'_{k,n_0}. In this case, we have

$$\ell(p_{k+1}) - \ell(p_k) = \ell(\tilde{p}) - \sqrt{m'_{k,n_0}}\, \mathrm{B}^k_{n_0}, \tag{25}$$

where $\tilde{p} := p_{k+1} \cap \mathrm{cut}(\Delta^k_{\tau_{k+1}}, V_{k+1})$ is the image of $]\!]u^k_{n_0}, v^k_{n_0}[\![\subset \Delta^k_{\tau_{k+1}}$ in $\mathrm{cut}(\Delta^k_{\tau_{k+1}}, V_{k+1})$. Observe that the root behaves as a uniform point in our cutting procedure, and that the rescaled tree $(m'_{k,n_0})^{-1/2}\Delta^k_{\tau_{k+1}}$ is a standard Brownian CRT. Thus, the induction hypothesis for $k = 1$ applies and with probability one there exists a sequence $(\tilde{\mathrm{T}}'_n, 0 \le n \le \tilde{N}')$, which is a sub-collection of the Δ^{k+1}_t, $t \in \mathcal{C}^{k+1}$, which are subsets of $\mathrm{cut}(\Delta^k_{\tau_{k+1}}, V_{k+1})$ (see Figure 3), rearranged in the decreasing order of their masses such that

$$\ell(\tilde{p}) = \sum_{n=0}^{\tilde{N}'} \sqrt{\mu(\tilde{\mathrm{T}}'_n)}\, \tilde{\mathrm{R}}'_n, \tag{26}$$

where $\mu(\tilde{\mathrm{T}}'_n)^{1/2}\tilde{\mathrm{R}}'_n$ is either the distance between two uniform independent points of $\tilde{\mathrm{T}}'_n$ or the distance between a uniform point and the root of $\tilde{\mathrm{T}}'_n$, so that $(\tilde{\mathrm{R}}'_n)$ is an i.i.d. family of Rayleigh distributed random variables. Furthermore, $(\mu(\tilde{\mathrm{T}}'_n)/m'_{k,n_0}, 0 \le n \le \tilde{N}')$ is an independent copy of $\mathbf{m}_1$. Then we set $\mathbf{F}'_{k+1} = (\mathrm{T}'_{k+1,n}, 0 \le n \le N'_{k+1})$ to be the rearrangement of the collection

$$\{\mathrm{T}'_{k,n} : 0 \le n \le N'_k, n \ne n_0\} \cup \{\tilde{\mathrm{T}}'_n : 0 \le n \le \tilde{N}'\}$$

such that $\mathbf{m}'_{k+1} := (\mu(\mathrm{T}'_{k+1,n}), 0 \le n \le N'_{k+1})$ is non-increasing. Inserting (25) and (26) into (24) yields the representation in (23) for $k+1$, which completes the proof of the induction step. □

As $(\mathbf{m}'_k)_{k\ge1}$ has the same distribution as $(\mathbf{m}_k)_{k\ge1}$, Lemma 13 also holds for $(\mathbf{m}'_k)_{k\ge1}$. Furthermore the concentration arguments already used in the course of the proof of Lemma 14 imply that, a.s.,

$$D - \mathbb{E}[D \,|\, \mathbf{m}'_k] = D - \sqrt{\pi/2}\sum_{n=0}^{N'_k}(m'_{k,n})^{1/2} = \sum_{n=0}^{N'_k}(m'_{k,n})^{1/2}\cdot\left(\mathrm{B}^k_n - \mathbb{E}[\mathrm{B}^k_n]\right) \to 0.$$

Since D does not vary with k, this implies that $(\mathbb{E}[D \,|\, \mathbf{m}'_k])_{k\ge1}$ converges almost surely. Since the sequence $(\mathbb{E}[\gamma_k(1,2) \,|\, \mathbf{m}_k])_{k\ge1}$ has the same distribution, it also converges almost surely and the proof of Lemma 15 is complete.

4 Direct construction of the complete reversal $\mathrm{shuff}(\mathcal{H})$

In this section, we finally prove that the operation which we described in the introduction as the dual to the complete cutting procedure makes sense, and is indeed the desired dual. This reduces to make the link between the collection of random variables $(A_x, x \in \mathrm{Br}(\mathcal{T}))$ and the iterative reversal of paths to the random leaves $(U_i)_{i\ge1}$. We prove here that the sequence of random leaves can be constructed as a measurable function of the family $(A_x, x \in \mathrm{Br}(\mathcal{T}))$.

4.1 Construction of one consistent leaf

Recall that $\mathcal{H}$ is a Brownian CRT rooted at ρ and with mass measure ν. Let $\{A_x, x \in \mathrm{Br}(\mathcal{H})\}$ be a family of independent random variables such that for every x, the point $A_x \in \mathrm{Sub}(\mathcal{H}, x)$ is chosen according to the restriction of the mass measure ν to $\mathrm{Sub}(\mathcal{H}, x)$.

One of the main constraints for the family $(U_i)_{i\ge1}$ is that the path to U_1 should be the first path to be reversed, and because of this, all the branch points x on the path between the root and U_1 should have choices A_x which are consistent with the reversing of the path to U_1 in the sense that for every branch point x on $[\![\rho, U_1]\!]$, one should have

$$\mathrm{Fr}(\mathcal{H}, x, A_x) = \mathrm{Fr}(\mathcal{H}, x, U_1).$$

Think of the discrete setting: if we have a rooted tree T and $(A_u, u \in T)$ a sequence of independent nodes such that A_u is distributed uniformly in the tree above u, then it is easy to construct a uniformly random node U_1 such that for every node u on the path between U_1 and the root, $[\![u, U_1]\!] \cap [\![u, A_u]\!] \setminus \{u\} \neq \varnothing$, or equivalently the points U_1 and A_u both lie in the same subtree of T rooted at one of the children of u. To do this, one simply needs to build the path from the root to U_1 by iteratively adding nodes as follows: start from the root; at some step where the current node is v, if $A_v \neq v$ then move to the first node in the direction to A_v, otherwise $A_v = v$ and set $U_1 = v$. The node U_1 is constructed so that the choices on the path $[\![\rho, U_1]\!]$ are consistent, and one easily verifies that U_1 is a uniformly random node. The idea is to adapt this technique to the continuous setting. To this aim, it suffices to verify that, when constructing a consistent path from the root, we make positive progress.

Let $v \in \mathcal{H} \setminus \{\rho\}$. We say that a branch point $x \in [\![\rho, v]\!] \cap \mathrm{Br}(\mathcal{H})$ is a *turning point* between the root and v if $A_x \in \mathrm{Fr}(\mathcal{H}, x, v)$. We denote by $\mathcal{N}_v$ the set of those x which are turning points between the root and v.

Lemma 17. *Let $\mathcal{H}$ be the Brownian CRT. For each $y \in \mathrm{Sk}(\mathcal{H})$, with probability one, $\mathcal{N}_y$ has at most finitely many elements.*

Proof. Let us write $\nu_y = \nu(\mathrm{Sub}(\mathcal{H}, y))$. As $\mathcal{H}$ is a Brownian CRT $\nu_y > 0$ almost surely for every $y \in \mathrm{Sk}(\mathcal{H})$. Let us write $\mathbb{P}_{\mathcal{H}}$ for the probability measure conditionally on $\mathcal{H}$. Since A_x is distributed according to the restriction of ν to $\mathrm{Sub}(\mathcal{H}, x)$, we have for each $x \in [\![\rho, y]\!] \cap \mathrm{Br}(\mathcal{H})$,

$$\mathbb{P}_{\mathcal{H}}(x \in \mathcal{N}_y) = \frac{\nu(\mathrm{Fr}(\mathcal{H}, x, y))}{\nu(\mathrm{Sub}(\mathcal{H}, x))} \leq \frac{\nu(\mathrm{Fr}(\mathcal{H}, x, y))}{\nu_y}$$

Then, we have

$$\begin{aligned}\mathbb{E}_{\mathcal{H}}[\mathrm{Card}(\mathcal{N}_y)] &= \sum_{x \in [\![\rho, y]\!] \cap \mathrm{Br}(\mathcal{H})} \mathbb{P}_{\mathcal{H}}(x \in \mathcal{N}_y) \\ &\leq \sum_{x \in [\![\rho, y]\!] \cap \mathrm{Br}(\mathcal{H})} \frac{\nu(\mathrm{Fr}(\mathcal{H}, x, y))}{\nu_y} = \frac{1}{\nu_y} < \infty, \quad \text{almost surely,}\end{aligned}$$

since the different fringe trees are disjoint. This shows that almost surely $\mathrm{Card}(\mathcal{N}_y) < \infty$. □

Now let us consider the following iterative process which constructs refining approximations Y_k, $k \geq 0$, of U. Start with $Y_0 = \rho$ and Z_0, which is a leaf of distribution ν. Supposing now that we have defined Y_k and Z_k, for some $k \geq 0$, if $\mathcal{N}_{Z_k} = \varnothing$, then we stop the process and set $U = Z_k$. Otherwise, there must exist some $y \in [\![\rho, Z_k]\!] \cap \mathrm{Sk}(\mathcal{H})$ such that $\mathcal{N}_y \neq \varnothing$. But by Lemma 17, $\mathcal{N}_y$ is finite a.s. so that there is an $x_0 \in \mathcal{N}_y$ which is closest to the root. Then we set $Y_{k+1} = x_0$ and $Z_{k+1} = A_{Y_{k+1}}$. Note that $Y_k \notin \mathcal{N}_{Z_k}$, so $Y_{k+1} \in \mathrm{Sub}(\mathcal{H}, Y_k) \setminus \{Y_k\}$.

Lemma 18. *As $k \to \infty$, we have $Y_k, Z_k \to U$ almost surely. Furthermore, U is a ν-distributed leaf.*

Proof. Let us first show that Z_k converges to some point U. If the process has stopped at some finite time, this is obvious. Otherwise, note that $(\mathrm{Sub}(\mathcal{H}, Y_k))_{k \geq 1}$ is a decreasing sequence of sets and define U in $\cap_{k \geq 0} \mathrm{Sub}(\mathcal{H}, Y_k)$ to be the point which minimizes the distance to the root. (So $\cap_{k \geq 0} \mathrm{Sub}(\mathcal{H}, Y_k) = \mathrm{Sub}(\mathcal{H}, U)$ and $Y_k \to U$.) Now, we claim that U is a leaf, so that $Z_k \to U$ as $k \to \infty$. To see this, suppose for a contradiction that U is not a leaf. Then $U \in \mathrm{Sk}(\mathcal{H})$ and with probability one, $\nu^\star := \nu(\mathrm{Sub}(\mathcal{H}, U)) > 0$. Note that by construction, we have $Z_k \notin \mathrm{Fr}(\mathcal{H}, Y_{k+1}, Z_k) \supset \mathrm{Sub}(\mathcal{H}, Y_{k+2})$, thus $Z_k \notin \mathrm{Sub}(\mathcal{H}, U)$ for each $k \geq 1$. However, for every k,

$$\begin{aligned}\mathbb{P}_{\mathcal{H}}(Z_i \notin \mathrm{Sub}(\mathcal{H}, U), i = 0, \ldots, k) &\leq \prod_{i=0}^{k} \left(1 - \frac{\nu(\mathrm{Sub}(\mathcal{H}, U))}{\nu(\mathcal{H}, Y_k)}\right) \\ &\leq (1 - \nu^\star)^k \to 0,\end{aligned}$$

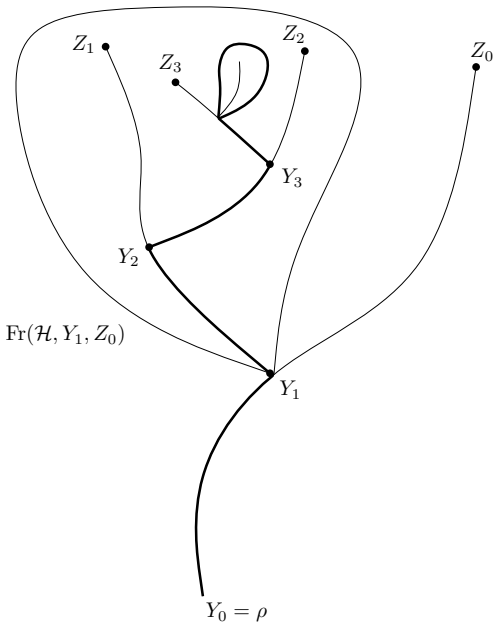

Figure 5: *The iterative construction of the path $Y_1, Y_2, \dots$ to the random leaf U.*

as $k \to \infty$. As a consequence, almost surely $\nu^\star = \nu(\mathrm{Sub}(\mathcal{H}, U)) = 0$ so that U is a.s. a leaf.

It now remains to prove that U is indeed ν-distributed. For this, it suffices to show that for every $x \in \mathcal{H}$, $\mathbb{P}_\mathcal{H}(U \in \mathrm{Sub}(\mathcal{H}, x)) = \nu(\mathrm{Sub}(\mathcal{H}, x))$. Note that since U is a leaf, we have $Z_k \to U$ as $k \to \infty$. We claim that for every $k \geq 0$, the leaf Z_k is ν distributed. Clearly, this would complete the proof. We proceed by induction on $k \geq 0$. For $k = 0$, Z_0 has distribution μ and the result is immediate. It will be useful to prove also the result for $k = 1$.

For a point $s \in [\![Y_0, Z_0]\!]$, we write $F_s = \mathrm{Fr}(\mathcal{H}, s, Z_0)$. For any $x \in [\![Y_0, Z_0]\!]$, since the branch points are countable, we have

$$\mathbb{P}_\mathcal{H}(Y_1 \in \mathrm{Sub}(\mathcal{H}, x) \,|\, Z_0, Y_0) = \prod_{s \in [\![Y_0, x[\![} \frac{\nu(\mathrm{Sub}(\mathcal{H}, s) \setminus F_s)}{\nu(\mathrm{Sub}(\mathcal{H}, s))} = \prod_{s \in [\![Y_0, x[\![} \left(1 - \frac{\nu(F_s)}{\nu(\mathrm{Sub}(\mathcal{H}, s))}\right), \quad (27)$$

since the choices of all the points are independent. There are only finitely $s \in [\![Y_0, x[\![$ for which $\nu(F_s) > \nu(\mathrm{Sub}(\mathcal{H}, x))/2$. For the others, $\nu(F_s)/\nu(\mathrm{Sub}(\mathcal{H}, s)) \leq 1/2$ and

$$1 - \frac{\nu(F_s)}{\nu(\mathrm{Sub}(\mathcal{H}, s))} \geq \exp\left(-2\frac{\nu(F_s)}{\mathrm{Sub}(\mathcal{H}, x))}\right).$$

It follows that the infinite product in (27) is absolutely convergent since $\sum_{s \in [\![Y_0, x]\!]} \nu(F_s) \leq 1$. Therefore, we have $\mathbb{P}_\mathcal{H}(Y_1 \in \mathrm{Sub}(\mathcal{H}, x) \,|\, Z_0, Y_0) = \nu(\mathrm{Sub}(\mathcal{H}, x))$. It follows that $\mathbb{P}_\mathcal{H}(Y_1 = s \,|\, Z_0, Y_0) = \nu(F_s)$, for $s \in [\![Y_0, Z_0]\!]$. Note that from our definition of F_s, this is indeed a probability distribution since $\sum_{s \in [\![Y_0, Z_0]\!]} \nu(F_s) = 1$. Now, for any $z \in \mathcal{H}$, we have

$$\mathbb{P}_\mathcal{H}(Z_1 \in \mathrm{Sub}(\mathcal{H}, z) \,|\, Y_1 = x, Z_0) = \frac{\nu(\mathrm{Sub}(\mathcal{H}, z))}{\nu(F_x)} \mathbf{1}_{\{z \in F_x\}},$$

which implies that, almost surely,

$$\mathbb{P}_\mathcal{H}(Z_1 \in \mathrm{Sub}(\mathcal{H}, z) \,|\, Z_0) = \sum_{x \in [\![Y_0, Z_0]\!]} \frac{\nu(\mathrm{Sub}(\mathcal{H}, z))}{\nu(F_x)} \mathbf{1}_{\{z \in F_x\}} \nu(F_x) = \nu(\mathrm{Sub}(\mathcal{H}, z)),$$

so that Z_1 has distribution ν. For the induction step, suppose now that Z_k has distribution ν. Conditionally on Y_k and Z_{k-1}, the point Z_k is distributed according to the restriction of ν to the set $\mathrm{Fr}(\mathcal{H}, Y_k, Z_{k-1})$.

Applying the result for $k = 1$ to the subtree $\mathrm{Fr}(\mathcal{H}, Y_k, Z_{k-1})$, we see that Z_{k+1} is also distributed according to the restriction of ν to $\mathrm{Fr}(\mathcal{H}, Y_k, Z_{k-1})$. So, Z_{k+1} and Z_k have the same conditional distribution, and it follows that Z_{k+1} is ν-distributed. Finally, since $Z_k \to U$ almost surely and for every k, Z_k is ν-distributed, it follows that U is a ν-distributed leaf and the proof is complete. □

Finally, we prove that U does not contain any of the "auxiliary" randomness used for the construction in the following sense:

Lemma 19. *Conditionally on $\mathcal{H}$, the random leaf U is a measurable function of $(A_x, x \in \mathrm{Br}(\mathcal{H}))$.*

Proof. It is clear from the construction, that U is a measurable function of $(A_x, x \in \mathrm{Br}(\mathcal{H}))$ and Z_0. It thus suffices to show that U is independent of Z_0. To see this, consider an independent copy Z_0' of Z_0 and let $(Y_k', Z_k')_{k\geq 1}$ be the sequence of random variables obtained from this initial choice. By Lemma 18, Z_k' converges almost surely to a leaf that we denote by U'. We now show that a.s. $U' = U$.

In this direction, we prove by induction that for every $k \geq 1$ we have $Y_k' \in [\![\rho, U]\!]$ and $Z_k' \in \mathrm{Fr}(\mathcal{H}, Y_k', U)$. To see that this is the case, observe that with probability one $Z_0' \wedge U \in \mathrm{Sk}(\mathcal{H})$ is a turning point for Z_0'. It is also the closest such point from the root and thus $Y_1' = Z_0' \wedge U$. Moreover, since $(Y_k)_{k\geq 0}$ is a path to U that is consistent with $(A_x, x \in \mathrm{Br}(\mathcal{H}))$, by construction $U \in \mathrm{Sub}(\mathcal{H}, Y_1') \setminus \mathrm{Fr}(\mathcal{H}, Y_1', Z_1')$. In other words, $Z_1' \in \mathrm{Fr}(\mathcal{H}, Y_1', U)$. Supposing now that the claim holds for all integers up to $k \geq 1$, we see that $Z_k' \wedge U \in]\!]Y_k', U[\![$ since $\mathcal{H}$ is a.s. binary. Again, $Z_k' \wedge U$ is a turning point, and there is no other such point on $]\!]Y_k', Z_k' \wedge U[\![$ so that $Y_{k+1}' = Z_k' \wedge U$. As before, $Y_{k+1}' \in [\![Y_r, Y_{r+1}]\!]$ for some $r \geq 0$, and because the path to U is consistent, it must be the case that $Z_k' \in \mathrm{Fr}(\mathcal{H}, Y_{k+1}', U)$.

Finally, recall that we proved in Lemma 18 that Y_k' also almost surely converges to U'. Since $U \in \mathrm{Sub}(\mathcal{H}, Y_k')$ for each $k \geq 1$, we have $U = U'$ and the proof is complete. □

4.2 The direct shuffle as the limit of k-reversals

It is now easy to use Lemma 18 in order to construct a sequence of i.i.d. leaves $(U_i)_{i\geq 1}$ which are distributed according to the mass measure ν, and is consistent with the collection $(A_x, x \in \mathrm{Br}(\mathcal{H}))$. We proceed inductively as follows. First set U_1 to be the ν-distributed leaf whose existence is guaranteed by Lemma 18. Then, assume that we have defined $(U_i)_{1\leq i\leq k}$ and set $\mathcal{S}_k = \cup_{1\leq i\leq k} [\![\rho, U_i]\!]$. Let U_{k+1}° be an independent ν-leaf in $\mathcal{H}$. With probability one $U_{k+1}^\circ \notin \mathcal{S}_k$, so $R_{k+1} := \{s \in \mathcal{H} : [\![s, U_{k+1}^\circ]\!] \cap \mathcal{S}_k = \varnothing\}$ is a non-empty subtree of $\mathcal{H}$ that also has positive mass. Then U_{k+1}° is distributed according to the restriction of ν to R_{k+1}, but may not be consistent with $(A_x, x \in \mathrm{Br}(R_{k+1}))$ in that subtree. By Lemma 18, there exists a random leaf U_{k+1}, distributed according to μ restricted to R_{k+1}, and which is consistent with the collection $(A_x, x \in \mathrm{Br}(R_{k+1}))$. One easily verifies that the collection $(U_i)_{i\geq 1}$ has the required properties.

Finally, the following result justifies the definition of $\mathrm{shuff}(\mathcal{H})$ that we gave in the introduction.

Proposition 20. *Let $\mathcal{H}$ be a Brownian CRT with mass measure ν. Let $(A_x, x \in \mathrm{Br}(\mathcal{H}))$ be a family of independent random variables such that A_x is distributed according to the restriction of ν to $\mathrm{Sub}(\mathcal{H}, x)$. Let $(U_i)_{i\geq 1}$ be a family of random leaves consistent with $(A_x, x \in \mathrm{Br}(\mathcal{H}))$. Then, the sequence*

$$(\mathrm{shuff}(\mathcal{H}; U_1, U_2, \ldots, U_k))_{k\geq 1}$$

converges a.s. as $k \to \infty$ in the sense of Gromov–Prokhorov, and we define $\mathrm{shuff}(\mathcal{H})$ to be the limit.

Note that this definition is indeed consistent with the algorithm given in the introduction: for every fixed $k \geq 1$, the branch points on $\mathrm{Span}(\mathcal{H}; U_1, \ldots, U_k)$ are used to form $\mathrm{shuff}(\mathcal{H}, U_1, \ldots, U_k)$; furthermore, since every branch point $x \in \mathrm{Br}(\mathcal{H})$ is in some $\mathrm{Span}(\mathcal{H}; U_1, \ldots, U_k)$ for k large enough, all the branch points are used to form the limit of $(\mathrm{shuff}(\mathcal{H}, U_1, \ldots, U_k))_{k\geq 1}$.

References

[1] L. Addario-Berry, N. Broutin, and C. Holmgren. Cutting down trees with a Markov chainsaw. *The Annals of Applied Probability*, 2013. (To appear).

[2] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[3] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[4] D. Aldous and J. Pitman. Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Related Fields*, 118(4):455–482, 2000.

[5] E. Baur and J. Bertoin. Cutting edges at random in large recursive trees. hal-00982497, 2014.

[6] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006. ISBN 978-0-521-86728-3; 0-521-86728-2.

[7] J. Bertoin. Fires on trees. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(4):909–921, 2012.

[8] J. Bertoin. The cut-tree of large recursive trees. hal-00859385, 2013.

[9] J. Bertoin and A. Gnedin. Asymptotic laws for nonconservative self-similar fragmentations. *Electronic Journal of Probability*, 9:575–593, 2004.

[10] J. Bertoin and G. Miermont. The cut-tree of large Galton-Watson trees and the Brownian CRT. *The Annals of Applied Probability*, 23:1469–1493, 2013.

[11] J.-M. Bismut. Last exit decompositions and regularity at the boundary of transition probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 69:65–98, 1985.

[12] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities - A nonasymptotic theory of independence*. Clarendon Press, Oxford, 2012.

[13] N. Broutin and M. Wang. Cutting down p-trees and inhomogeneous continuum random trees. arXiv:1408.0144 [math.PR], 2014.

[14] M. Drmota, A. Iksanov, M. Moehle, and U. Roesler. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures and Algorithms*, 34:319–336, 2009.

[15] S. N. Evans. *Probability and real trees*, volume 1920 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005.

[16] J. Fill, N. Kapur, and A. Panholzer. Destruction of very simple trees. *Algorithmica*, 46:345–366, 2006.

[17] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probability Theory and Related Fields*, 145(1-2):285–322, 2009.

[18] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, english edition, 2007.

[19] C. Holmgren. Random records and cuttings in binary search trees. *Combinatorics, Probability and Computing*, 19:391–424, 2010.

[20] C. Holmgren. A weakly 1-stable limiting distribution for the number of random records and cuttings in split trees. *Advances in Applied Probability*, 43:151–177, 2011.

[21] A. Iksanov and M. Möhle. A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electronic Communications in Probability*, 12:28–35, 2007.

[22] S. Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006. ISSN 1042-9832.

[23] J.-F. Le Gall. Random trees and applications. *Probability Surveys*, 2:245–311, 2005.

[24] A. Meir and J. Moon. Cutting down random trees. *Journal of the Australian Mathematical Society*, 11: 313–324, 1970.

[25] A. Panholzer. Cutting down very simple trees. *Quaestiones Mathematicae*, 29:211–228, 2006.

[26] M. Perman, J. Pitman, and M. Yor. Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields*, 92(1):21–39, 1992.

A Some facts about the Brownian CRT

In this section, we provide a proof of Lemma 12, which says that the sequence of (mass rescaled) real trees encountered on the path between two random points in the k-reveral are independent Brownian CRT. The proof is based on the scaling property of the Brownian excursion, Bismut's path decomposition of a Brownian excursion [11], and the size-biased sampling of a Poisson point process due to Perman et al. [26]. For the sake of precision, the arguments are phrased in terms of excursions rather than real trees, but we make sure to give the tree-based intuition as well.

BISMUT'S DECOMPOSITION. Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion. Let $\mathbf{N}$ be the Itô measure for the excursions of $|B|$ away from 0, which is a σ-finite measure on the space of non-negative continuous paths $\mathscr{C}(\mathbb{R}_+, \mathbb{R}_+)$. Let $w = (w_s)_{s\geq 0}$ be the coordinate process. In particular, if we denote by $\zeta := \inf\{s > 0 : w_s = 0\}$ the lifetime of an excursion, then $\mathbf{N}(\zeta \in dr) = dr/\sqrt{2\pi r^3}$. We denote by $\mathbf{N}^{(1)}$ the law of the normalized excursion. For $r > 0$, we set $\mathbf{N}^{(r)}$ to be the distribution of the rescaled process $(\sqrt{r}w_{s/r})_{0\leq s\leq r}$ under $\mathbf{N}^{(1)}$. Then it follows from the scaling property of Brownian motion that $\mathbf{N}^{(r)}$ is the law of w under $\mathbf{N}(\,\cdot\,|\,\zeta = r)$. We have

$$\mathbf{N}(\,\cdot\,) = \int \mathbf{N}(\zeta \in dr)\mathbf{N}^{(r)}(\,\cdot\,). \tag{28}$$

If we set $\mathscr{S}(w) \in \mathscr{C}(\mathbb{R}_+, \mathbb{R}_+)$ to be the process $(\mathscr{S}(w))(s) = \zeta^{-1/2}w_{s\zeta}$ for $s \geq 0$, then it follows from (28) that, under $\mathbf{N}$, $\mathscr{S}(w)$ is independent of ζ and is distributed as the normalized excursion.

Let $(Z_k)_{k\geq 0}$ be a sequence of independent variables uniformly distributed on $[0, \zeta]$. If we see w as encoding a Brownian CRT, then $(Z_k)_{k\geq 0}$ is a sequence of leaves sampled according to the mass measure. The CRT is decomposed along the path leading to the leaf corresponding to Z_0: set $K := w_{Z_0}$ and let

$$\overleftarrow{w}^\circ(t) = w_{(Z_0-t)\vee 0}; \quad \overrightarrow{w}^\circ(t) = w_{t+Z_0}, \quad t \geq 0.$$

We need the following notation to describe precisely the spinal decomposition. Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function with compact support and suppose that $h(0) > 0$. Define $\underline{h}(s) = \inf_{0\leq u\leq s} h(u)$. Let $\{(l_i, r_i), i \in \mathcal{I}(h)\}$ be the excursion intervals of $h - \underline{h}$ away from 0, which are the connected components of the open set $\{s \in \mathbb{R}_+ : h(s) - \underline{h}(s) > 0\}$. For each $i \in \mathcal{I}(h)$, let

$$h^i(s) = \big(h - \underline{h}\big)\big((l_i + s) \wedge r_i\big), \quad s \geq 0$$

be the excursion of $h - \underline{h}$ over the interval (l_i, r_i). We denote by $\mathcal{P}(h)$ the point measure on $\mathbb{R}_+ \times \mathscr{C}(\mathbb{R}_+, \mathbb{R}_+)$ defined by

$$\mathcal{P}(h) = \sum_{i\in\mathcal{I}(h)} \delta_{(h(0)-h(l_i), h^i)}.$$

We set $\mathcal{Q}$ to be the sum of the point measures $\mathcal{P}(\overrightarrow{w}^\circ)$ and $\mathcal{P}(\overleftarrow{w}^\circ)$ on $\mathbb{R}_+ \times \mathscr{C}(\mathbb{R}_+, \mathbb{R}_+)$:

$$\mathcal{Q} = \mathcal{P}(\overrightarrow{w}^\circ) + \mathcal{P}(\overleftarrow{w}^\circ) := \sum_{j\geq 1} \delta_{(s_j, w^j)},$$

where the last expression above serves as the definition of s_j and w^j, for $j \geq 1$.

Lemma 21 (Bismut's decomposition, [11]). *For each $m > 0$, $\mathcal{Q}$ under $\mathbf{N}(\,\cdot\,|K = m)$ has the distribution of a Poisson point measure of intensity measure $\mathbf{1}_{[0,m]}2dt \otimes \mathbf{N}$.*

SIZE-BIASED ORDERING OF THE EXCURSIONS. For each $j \geq 1$, we denote by I_j the excursion interval (l_i, r_i) associated with w^j. We write ζ_j for the length of I_j. Notice that $\mathbf{N}$-a.s. $\sum_{j\geq 1} \zeta_j = \zeta$. Let $(\kappa_i)_{i\geq 1}$ be the permutation of $\mathbb{N}$ induced by $(Z_n)_{n\geq 1}$ as follows. Let κ_1 be the index such that $Z_1 \in I_{\kappa_1}$, then

κ_1 is well-defined almost surely. For each $i \geq 1$, let $\sigma(i) := \inf\{n \geq 1 : Z_n \notin \cup_{1\leq n\leq i} I_{\kappa_i}\}$, then let κ_{i+1} be the index such that $Z_{\sigma(i)} \in I_{\kappa_{i+1}}$. Let us point out that according to Lemma 21, $(\zeta_j)_{j\geq 1}$ under $\mathbf{N}(\cdot \,|\, K = m)$ is distributed as the jumps of a $1/2$-stable subordinator before time $2m$, and $(\zeta_{\kappa_i})_{i\geq 1}$ is the size-biased sampling introduced in [26]. Then by Palm's formula and the product form of the intensity measure in Lemma 21, we have

Lemma 22. *Under $\mathbf{N}^{(1)}(\cdot \,|\, K = m)$, the three sequences $(\zeta_{\kappa_i})_{i\geq 1}$, $(s_{\kappa_i})_{i\geq 1}$, $(\mathscr{S}(w^{\kappa_i}))_{i\geq 1}$ are independent. Moreover, $(\zeta_{\kappa_i})_{i\geq 1}$ is a Markov chain, $(s_{\kappa_i})_{i\geq 1}$ is an i.i.d. sequence of uniform variables on $[0, m]$, and $(\mathscr{S}(w^{\kappa_i}))_{i\geq 1}$ is an i.i.d. sequence of common law $\mathbf{N}^{(1)}$.*

Proof. Let $f, g, h : \mathbb{R}_+ \to \mathbb{R}$ be some continuous bounded functions, let $H : \mathscr{C}(\mathbb{R}_+, \mathbb{R}_+) \to \mathbb{R}$ and $G : \mathbb{R}_+ \times \mathscr{C}(\mathbb{R}_+, \mathbb{R}_+) \to \mathbb{R}$ be continuous and bounded. We use the notations:

$$\mathcal{Q}^- = \mathcal{Q} - \delta_{(s_{\kappa_1}, w^{\kappa_1})} \qquad \text{and} \qquad \mathcal{Q}^{j-} = \mathcal{Q} - \delta_{(s_j, w^j)}, \quad j \geq 1.$$

Note that $\zeta = \sum_{j\geq 1} \zeta_j$ is distributed as X_{2m}, where $(X_s)_{s\geq 0}$ is a $1/2$-stable subordinator. We denote by θ the density of X_{2m}. Then, by Palm's formula,

$$\begin{aligned}
&\mathbb{E}\left[f(s_{\kappa_1}) g(\zeta_{\kappa_1}) H(w^{\kappa_1}) G(\mathcal{Q}^-) h(\zeta)\right] \\
&= \mathbb{E}\left[\sum_{j\geq 1} \frac{\zeta_j}{\zeta} f(s_j) g(\zeta_j) H(w^j) G(\mathcal{Q}^{j-}) h(\zeta)\right] \\
&= \int_0^\infty dz \theta(z) h(r+z) \int_0^m 2dt g(t) \int_0^\infty \frac{dr}{\sqrt{2\pi r^3}} f(r) \frac{r}{r+z} \mathbf{N}^{(r)}\big(H(w^j)\big) \mathbb{E}[G(\mathcal{Q}) \,|\, \zeta = z].
\end{aligned}$$

It follows that, for $0 < r < 1$, and $0 < t < m$, we have

$$\mathbb{P}\left(\zeta_{\kappa_1} \in dr, s_{\kappa_1} \in dt, \mathscr{S}(w^{\kappa_1}) \in dw \,\middle|\, \zeta = 1\right) = \frac{\theta(1-r)dr}{\Lambda\sqrt{2\pi r}} \cdot \frac{dt}{m} \cdot \mathbf{N}^{(1)}(dw)$$

where $\Lambda = \int_0^1 dr \theta(1-r)/\sqrt{2\pi r}$, and given $\zeta_{\kappa_1} = r$, $\mathcal{Q}^-$ is independent of $(s_{\kappa_1}, \mathscr{S}(w^{\kappa_1}))$ and has distribution $\mathbf{N}(\cdot \,|\, \zeta = 1 - r, K = m)$. From there, a simple induction argument yields the claim. □

THE TREES ALONG THE PATH BETWEEN TWO RANDOM POINTS. We are finally in position of proving Lemma 12. Let $(\pi_j)_{j\geq 1}$ be the subsequence of $(\kappa_i)_{i\geq 1}$ defined by

$$\pi_1 = \kappa_1, \qquad \text{and if } \tau_j = \inf\{i > \pi_j : s_{\kappa_i} > s_{\pi_j}\} \text{ then } \pi_{j+1} = \kappa_{\tau_j}, \quad j \geq 1.$$

Then by definition, $(s_{\pi_j}, j \geq 1)$ has the same distribution as $(d(\rho, x_1(m)), m \geq 0)$, where $(x_1(m), m \geq 0)$ is the sequence defined in Proposition 6. Observe that $(\pi_j)_{j\geq 1}$ depends only on $(s_{\kappa_i})_{i\geq 1}$, which is independent of $(\mathscr{S}(w^{\kappa_i}))_{i\geq 1}$ according to the previous lemma. We deduce that

Lemma 23. *Under $\mathbf{N}^{(1)}(\cdot \,|\, K = m)$, the three sequences $(\zeta_{\pi_j})_{j\geq 1}$, $(s_{\pi_j})_{j\geq 1}$, and $(\mathscr{S}(w^{\pi_j}))_{j\geq 1}$ are independent. Moreover, $(\mathscr{S}(w^{\pi_j}), j \geq 1)$ is an i.i.d. sequence of common law $\mathbf{N}^{(1)}$.*

Finally, let $(Z'_k)_{k\geq 1}$ be another sequence of i.i.d. random variables uniformly distributed on $[0, \zeta]$, independent of $(Z_k)_{k\geq 1}$. Define $(\kappa'_i)_{i\geq 1}$ and $(\pi'_j)_{j\geq 1}$ using $(Z'_k)_{k\geq 1}$ in the same way that $(\kappa_i)_{i\geq 1}$ and $(\pi_j)_{j\geq 1}$ were defined using $(Z_k)_{k\geq 1}$. Let

$$\mathscr{I} := \inf\{j \geq 1 : \exists k \geq 1 \text{ such that } s_{\kappa_j} = s_{\kappa'_k}\}, \quad \mathscr{I}' := \inf\{k \geq 1 : \exists j \geq 1 \text{ such that } s_{\kappa'_k} = s_{\kappa_j}\}.$$

Note that for each pair (j_0, k_0), the event $\{\mathscr{I} = j_0, \mathscr{I}' = k_0\}$ depends only on $(\zeta_{\pi_j}, 1 \leq j \leq j_0)$ and $(\zeta_{\pi'_k}, 1 \leq k \leq k_0)$. Therefore, on this event, $\{\mathscr{S}(w^{\pi_j}), 1 \leq j \leq j_0\} \cup \{\mathscr{S}(w^{\pi_k}), 1 \leq k \leq k_0 - 1\}$ are $j_0 + k_0 - 1$ independent copies of w under $\mathbf{N}^{(1)}$. Integrating with respect to the law of K under $\mathbf{N}^{(1)}$ (the Rayleigh distribution) proves Lemma 12.