No Shannon effect induced by `And/Or` *trees* †

Antoine Genitrini[1,2] and Bernhard Gittenberger[3] and Cécile Mailler[4]

[1]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France.*
[2]*CNRS, UMR 7606, LIP6, F-75005, Paris, France. Email: Antoine.Genitrini@lip6.fr.*
[3]*Technische Universität Wien, Wiedner Hauptstrasse 8-10/104, A-1040 Wien, Austria.*
Email: gittenberger@dmg.tuwien.ac.at.
[4]*Department of Mathematical Sciences, University of Bath, BA2 7AY Bath, UK. Email: c.mailler@bath.ac.uk.*

Quantitative logic has been the subject of an increasing interest since a seminal paper by Chauvin et al. in 2004, which presented the first Analytic Combinatorics approach of the subject. Since then, the understanding of random Boolean trees has been deeply widened, even if the question of Shannon effect remains open for the majority of the models. We focus in this paper on the original case of Catalan `And/Or` trees and propose a new specification of those objects that implies easier ways to describe large families of trees. Equipped with this specification, we prove that the model of Catalan `And/Or` binary trees do not exhibit Shannon effect, i.e. there exists a family of functions with small complexities, that have a positive probability.

Keywords: Boolean functions; Probability distribution; Random Boolean formulas; Random trees; Asymptotic ratio; Analytic combinatorics.

1 Introduction

Pick up at random a Boolean formula of a given size: what is the distribution of the random Boolean function it represents? Defining models in which this question has a meaning and answering it in these models is the domain of quantitative logic. To our knowledge, the first attempts in this direction go back to the 90's, by Paris et al. [PVW94] and later Lefmann and Savický [LS97]. The framework they propose is to use the representation of Boolean formulas by plane binary trees and then rely on the good combinatorial properties of these trees. The domain has then strongly developed after the seminal contribution of Chauvin et al. [CFGG04], proposing the first Analytic Combinatorics approach of the question. Since 2004, many different models have been studied via Analytic Combinatorics, leading to a better understanding of random Boolean trees.

The first approaches [PVW94, LS97, CFGG04] focused on the `And/Or` logical system, i.e. on Boolean formulas built with the connectives AND and OR and with the negation. They pick up a binary plane Boolean tree uniformly at random among all trees with m leaves, and then let m tend to infinity: the induced distribution on the set of Boolean functions is called the *Catalan tree distribution*. Their initial

†This research was partially supported by the P.H.C. AMADEUS project 29281NE, the ÖAD project F03/2013 and the FWF grant SFB F50-03.

results have been deeply improved by Kozik [Koz08], via Analytic Combinatorics and a very powerful *ad hoc* machinery called the *pattern theory*. Equivalent results have been shown for the implicational system by Fournier et al. [FGGG12], but there a different approach was required. The theory has then been extended to many directions: for example by considering non plane and non binary trees in order to take into account the logical properties of the logical connectives (for example the commutativity and associativity of AND and OR) [GGKM12, GGKM]; or by considering different distributions on Boolean trees [CFGG04, CGM11].

In this paper, however, we aim at revisiting the original And/Or Catalan tree distribution in order to prove that this model does not exhibit the *Shannon effect*. Riordan and Shannon have proved, in the paper [RS42], that a uniformly distributed random Boolean function on n variables is almost surely of exponential *complexity*: we say that the uniform distribution on Boolean functions of fixed size exhibits the Shannon effect. We prove in the present paper that the Catalan tree distribution does not. To prove this new result, we needed to develop a new framework: a new way to specify And/Or trees and thus a new way to count them.

We would like to point out, that the study of the Shannon effect in random Boolean tree models is only done in very few models [GG10, GGKM12], both models based on the connective of IMPLICATION, in the case of plane or non-plane trees. There is thus a need to progress in this direction. Moreover, we think that it is important to find universal methods to study those models, i.e. methods that could be used for both the And/Or and the implicational logical systems (note that Kozik's pattern theory does not apply for implicational trees). This was successfully done in the case of branching processes as underlying tree model [GG10]. The new proof happens to be very similar to the ones used in the implicational system [GG10, GGKM12]. The method we use has some common base with the expansion tools, developped for example in [FGGG12], but it is very different in the way the tools are used.

This paper is organised in two sections. The first one, Section 2, is devoted to the definition of the Catalan tree model and to the introduction of our new point of view on And/Or trees. The second one, Section 3, states and prove our main result on Shannon effect in the Catalan tree model and its proof.

2 Description of the model and specifications

2.1 Context

Definition 1 *A* **Catalan tree** *is a rooted full binary plane tree (meaning that each node has either 0 or 2 children). We define the* **size** *of a Catalan tree as the number of its leaves. The combinatorial class of Catalan trees is denoted by $\mathcal{C}$. The corresponding ordinary generating function is $C(z) = \sum_{m>0} C_m z^m$, where C_m is the number of Catalan trees of size m.*

It is well known that $C(z) = \dfrac{1-\sqrt{1-4z}}{2}$ and $C_m = \dfrac{1}{m} \cdot \dbinom{2m-2}{m-1}$.

Definition 2 *An And/Or tree on n variables is a Catalan tree whose internal nodes are labelled by a connective from the set $\{\wedge, \vee\}$ (resp. AND and OR) and whose leaves are labelled by a literal from $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$, where $\bar{x}_i$ denotes the negation of x_i. The combinatorial class of And/Or trees on n variables is denoted by $\mathcal{T}_n$, the number of such trees of size m by $T_{m,n}$.*

Each And/Or tree is equivalent to a Boolean expression and thus *represents*, or *computes*, a Boolean function with n variables, i.e. a function from $\{0,1\}^n$ into $\{0,1\}$. An example is given in Fig. 1. We will from now denote the set of Boolean functions with n variables by $\mathcal{F}_n$.

Definition 3 *The* **complexity** $L(f)$ *of a non constant Boolean function* $f \in \mathcal{F}_n$ *is the size of the smallest trees calculating it. We call the trees realizing this complexity the* **minimal trees** *of* f*. The complexity of the constant functions* true : $(x_1, \ldots, x_n) \mapsto 1$ *and* false : $(x_1, \ldots, x_n) \mapsto 0$ *is defined to be 0.*

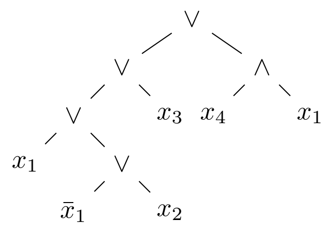

Fig. 1: The above tree is equivalent to the Boolean formula $([x_1 \vee (\bar{x}_1 \vee x_2)] \vee x_3) \vee (x_4 \wedge x_1)$ and represents the constant Boolean function $\mathtt{true} : (x_1, \ldots, x_n) \mapsto 1$.

The idea of quantitative logic is historically the following: Equip the set of `And/Or` trees of size m with the uniform distribution and denote by $\mu_{m,n}$ the induced distribution on the set of Boolean functions in n variables. Can we prove that the sequence $(\mu_{m,n})_{m\geq 1}$ converges? And what are the properties of the limiting distribution (if it exists)?

Since this model has already been studied in the literature, where it is sometimes called the **Catalan model**, we already know the following result:

Theorem 4 ([LS97, CFGG04]) *The sequence of distributions* $(\mu_{m,n})_{m\geq 1}$ *converges to a limiting distribution on the set of Boolean function with* n *variables. This distribution is denoted by* μ_n *and called the* **Catalan distribution***.*

Moreover, we have the following inequalities: for all Boolean functions $f \in \mathcal{F}_n$*,*

$$\frac{1}{4}\left(\frac{1}{8n}\right)^{L(f)+1} \leq \mu_n(f) \leq (1 + O(1/n))\mathrm{e}^{-c\frac{L(f)+1}{n^2}},$$

asymptotically when n *tends to* $+\infty$*.*

This is the best result, in full generality, up to now. Kozik [Koz08] has proven the following more precise result:

Theorem 5 ([Koz08]) *Let* $f = f(x_1, \ldots, x_{n_0})$ *be a Boolean function and* $\mu_n(f)$ *denote the proportion of all* `And/Or` *trees on* n *which represent* f *(*cf. *Definition 8). Then there exists a constant* $\lambda_f > 0$ *such that, for* n *tending to infinity,*

$$\mu_n(f) = \frac{\lambda_f}{n^{L(f)+1}} + O\left(\frac{1}{n^{L(f)+2}}\right).$$

This result suggests that the Catalan distribution *gives more weight to low complexity functions*, but this result is too weak to claim that the Catalan distribution *does not exhibit the Shannon effect*. One has to consider sequences of functions indexed on n.

Definition 6 *We say that a sequence of distributions $(\nu_n)_{n\geq 1}$, respectively defined on the sets $(\mathcal{F}_n)_{n\geq 1}$,* **exhibit the Shannon effect** *if, for all $\varepsilon > 0$,*

$$\nu_n\left(\left\{f \text{ such that } L(f) \geq \frac{2^n(1-\varepsilon)}{\log_2 n}\right\}\right) \to 1$$

when n tends to $+\infty$.

Remark: It is shown (see [Lup62] that the maximal complexity of a Boolean function with n variables is less than $\frac{2^n}{\log_2 n}(1 + o(1))$. Thus, the sequence $(\nu_n)_{n\geq 1}$ exhibit the Shannon effect if, asymptotically when n tends to $+\infty$, almost all Boolean functions (under ν_n) have almost maximal complexity.

To prove that a sequence of distributions $(\nu_n)_{n\geq 1}$ does not exhibit the Shannon effect, it is enough to find a sub-exponential function $g(n)$ and a constant $\alpha > 0$ such that, for all $n \geq 1$,

$$\nu_n\left(\{f \text{ such that } L(f) \leq g(n)\}\right) \geq \alpha > 0.$$

The approach is thus very different from the proof of Theorem 5 since one has to consider a set of Boolean functions, among them functions whose complexity does depend on n. This explains why the Shannon effect has only been studied in a few models: the implicational model [GG10], the general implicational model [GGKM12], and the models based on branching processes [GG10]. The successful approach in those two articles is based on a very careful Analytic Combinatorics study.

Corollary 7 (of Theorem 5) *If K is some constant which is independent of n. Then the set of all Boolean functions up to complexity K is not sufficiently large to refute the Shannon effect.*

Proof: The number of functions of complexity k is bounded by $\alpha \cdot n^k$ (where α is a constant), and therefore, by using Theorem 5, the probability of functions of complexity smaller or equal to k is $O(1/n)$. Thus, we need a class of functions of complexity up to some bound K_n tending to infinity as n tends to infinity to have a chance to prove that there is no Shannon effect. $\square$

Our aim in the present paper is to introduce a new specification for `And/Or` trees which simplifies the Analytic Combinatorics involved in their study and therefore allows us to tackle the Shannon effect question in this context.

2.2 New specification of `And/Or` trees

The historical specification of `And/Or` trees is the following [CFGG04]: there are $2n$ different trees of size one (one for each literal $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$), and an `And/Or` tree is either a tree of size one ($2n\cdot\mathcal{Z}$) or an internal node (two choices) with two subtrees from $\mathcal{T}$. Namely,

$$\mathcal{T} = 2\cdot(\mathcal{T}\times\mathcal{T}) + 2n\cdot\mathcal{Z}.$$

We propose here another point of view that simplifies the descriptions of the subfamilies of trees needed later on and therefore the technical calculations. Maybe, with this new description we could rederive the results of [Koz08] while circumventing the pattern theory but keeping the key ideas of the proofs.

Let us denote by $\mathcal{A}$ the family of `And/Or` trees that are rooted by an $\wedge$ label, and by $\mathcal{O}$ the family of trees that are rooted by an $\vee$ label. We have $\mathcal{T} = 2n\mathcal{Z} + \mathcal{A} + \mathcal{O}$, and we describe the $\wedge$-rooted trees as follows: it is a Catalan tree of $\wedge$-labels (of size at least two), whose leaves are substituted either by (Boolean decorated) leaves ($2n\mathcal{Z}$) or by $\vee$-rooted trees. Namely,

$$\mathcal{A} = (\mathcal{C} \setminus \mathcal{Z}) \circ (2n \cdot \mathcal{Z} + \mathcal{O}).$$

Thus, in terms of generating functions, we have via the symbolic method (see [FS09] for an introduction to this method)

$$A(z) = K(O(z) + 2nz), \tag{1}$$

where $K(z) = C(z) - z$ is the generating function of Catalan trees of size at least 2, $A(z)$ (resp. $O(z)$) is the generating function of $\wedge$-rooted trees (resp. $\vee$-rooted trees). By symmetry $A(z) = O(z)$, thus

$$A(z) = \frac{1 - 8nz - \sqrt{1 - 16nz}}{8}.$$

Obviously, we again obtain the classical generating enumerating all trees,

$$T(z) = 2nz + 2A(z) = \frac{1 - \sqrt{1 - 16nz}}{4},$$

which is directly translated from the classical specification: $\mathcal{T} = 2n \cdot \mathcal{Z} + 2 \cdot (\mathcal{T} \times \mathcal{T})$.

Observe that an `And/Or` tree can be seen in the following way: Instead of looking at it node by node, we partition the set of internal nodes into clusters of nodes carrying the same label. For example, if the root is labelled by $\vee$, then we take all nodes labelled by $\vee$ and connected to the root by an $\vee$-only path into a super-$\vee$-node. Repeat this procedure recursively. Then, we arrive at a tree consisting of such supernodes, which we will call $\wedge$- and $\vee$-**clusters** in the sequel, and leaves. This tree is not binary any more, and on each path the labels of the clusters are alternating. We will refer to this viewpoint as the **cluster tree** of a given `And/Or` tree. Equation (1) corresponds to this way of looking at an `And/Or` tree.

2.3 Distribution on `And/Or` trees in details

Let us recall the notions and notations which have become common in quantitative logics (cf. [FGGG12] for example).

Definition 8 *Let $\mathcal{X}$ be a subfamily of $\mathcal{T}$, we define its* **limiting ratio** $[\mathcal{X}]_n$ *as the following limit, if it exists:*

$$[\mathcal{X}]_n := \lim_{m \to +\infty} \frac{X_{m,n}}{T_{m,n}},$$

where $T_{m,n}$ is the number of `And/Or` *trees of size m and $X_{m,n}$ is the number of elements of $\mathcal{X}$ of size m.*

Remark: Let f be a Boolean function in $\mathcal{F}_n$, and let $\mathcal{X}(f)$ be the family of `And/Or` trees that compute f. Then $[\mathcal{X}(f)]_n$ exists and we set $\mu_n(f) = [\mathcal{X}(f)]_n$.

Fact 9 *Let $\mathcal{X}$ be a subfamily of $\mathcal{T}$. If the dominant singularity of the generating function $X(z) = \sum_{m \geq 0} X_{m,n} z^m$ of $\mathcal{X}$ is $\rho = 1/16n$ and if this singularity is a square-root singularity, then*

$$[\mathcal{X}]_n = \lim_{z \to \rho} \frac{X'(z)}{T'(z)}.$$

3 No Shannon effect in And/Or trees

Theorem 10 *The Catalan distribution does not exhibit the Shannon effect.*

Let us mention here a difference in the approach that we will use in this paper than the one used in [GG10]. The generating functions that appear there are not as intricate as the one we get in the And/Or tree model. So, in the implicational model the authors have tried to obtain the best possible lower bound for the probability of a well-chosen family of functions. In this paper, our goal is to prove that there is no Shannon effect. We will not care for the final value for the probability of a suitably chosen family of functions, but only prove that it is positive.

Definition 11 *Let t be an And/Or tree and ν_α be a leaf labelled by the literal α in t. Denote by η the root of the cluster containing ν_α. The subtree rooted by η is denoted u.*
Let ν be a node of u, and v the subtree with root ν.

- *If ν_α belongs to an $\vee$-cluster, then replace v by $(v \vee t_e)$ or $(t_e \vee v)$ in t, where t_e is a tree with a root labelled by $\wedge$ and such that at least one of the leaves that are attached to the topmost $\wedge$-cluster is labelled by α.*
- *If ν_α belongs to an $\wedge$-cluster, then replace v by $(v \wedge t_e)$ or $(t_e \wedge v)$ in t, where t_e is a tree with a root labelled by $\vee$ and such that at least one of the leaves that are attached to the topmost $\vee$-cluster is labelled by α.*

*The obtained tree $\tilde{t}$ is called an α-*expansion *of t.*

Remark: In every node of u we can do an expansion according to the label of ν_α.

Lemma 12 *If t is a tree computing a Boolean function f, then for all literals α, all α-expansions of t are trees computing f.*

Proof: A proof by case is relatively straightforward. Assume that ν_α belongs to an $\vee$-cluster. Then u computes a function of the form $\alpha \vee f$. If $\alpha = \texttt{True}$, then the value of f is irrelevant and thus an expansion can do no harm. If $\alpha = \texttt{False}$, then observe that t_e is rooted by an $\wedge$-cluster which has a leaf α. Thus t_e attains the value False and replacing v by $v \vee t_e$ has no effect, either. The other cases are analogous. □

To prove Theorem 10, we will enumerate a subfamily of the expansions of all trees of size at most $g(n)$, for all literals α. We will constrain the expansion-structures in order to be able to retrieve the node in an expansion where the grafting has been done. This avoids double counting. We constrain the shapes in the following way: the grafted tree t_e has a shape $\alpha \vee (t_1 \vee t_2)$ (resp. $\alpha \wedge (t_1 \wedge t_2)$) where t_1 and t_2 are two subtrees of size at least $g(n)$ (we denote by $G(z)$ the generating function of such trees). Thus, given an expansion, find the topmost node that has two subtrees of size at least $g(n)$; its parent is the root of the grafted subtree t_e, denoted by ν in Fig. 2.

Fix a literal α. Let $\hat{N}(z,u)$ (resp. $\check{N}(z,u)$) be the generating function of $\wedge$-rooted (resp. $\vee$-rooted) And/Or trees, where z marks the size and u marks the number of nodes allowing an α-expansion. It is given by

$$\check{N}(z,u) = K((2n-1)z + \hat{N}(z,u)) + \sum_{p \geq 1} \frac{u^{2p-1} z^p}{p!} K^{(p)}(A(u^2 z) + u^2(2n-1)z),$$

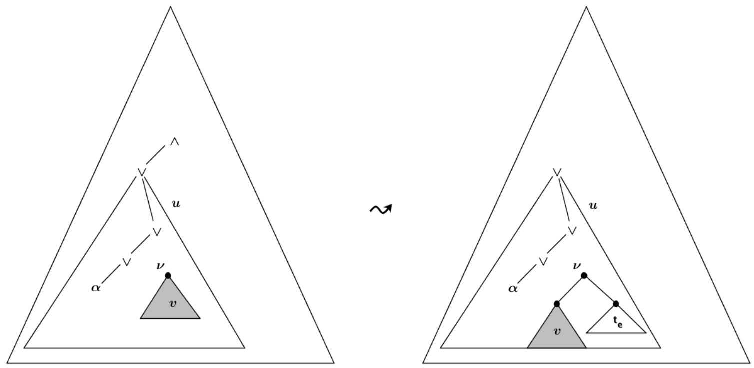

Fig. 2: On the right, the `And/Or` tree is an expansion from the tree represented on the left.

where

- the first term $K((2n-1)z + \hat{N}(z,u))$ stands for the case where there is no leaf of the topmost $\vee$-cluster that is labelled by α;
- the term $u^{2p-1}z^p K^{(p)}(A(u^2z)+u^2(2n-1)z)/p!$ stands for the case, where exactly p leaves of the topmost $\vee$-cluster are labelled by α: in this case, the choice of those p leaves yields the derivative and an expansion is possible at every node in the tree. Thus, a tree having k leaves contributes $z^k u^{2k-1}$ to the generating function $\check{N}(z,u)$.

By symmetry, we have an analogous equation for $\hat{N}(z,u)$, thus let us define $N(z,u) := \check{N}(z,u) = \hat{N}(z,u)$. Consequently we obtain:

$$N(z,u) = K((2n-1)z + N(z,u)) + \sum_{p\geq 1} \frac{u^{2p-1}z^p}{p!} K^{(p)}(A(u^2z) + u^2(2n-1)z), \tag{2}$$

Since we have avoided multiple-counting, we get the following inequality:

$$\mu_n(\{f \text{ such that } L(f) \leq g(n)\}) \geq \lim_{m\to+\infty} 4n \cdot \sum_{r=1}^{g(n)} [z^r]\partial_u N(z,1)\, \frac{[z^{m-r}]G(z)}{[z^m]T(z)}. \tag{3}$$

The factor $4n$ represents the choice of the literal α (with respect to which the expansions are done) and the label of the root of the tree; the sum is over all r from 1 to $g(n)$, where r represents the size of the original trees that are then expanded); the factor $[z^r]\partial_u N(z,1)$ counts the sum over all trees of size r of the different places where an α-expansion is possible; and the factor $[z^{m-r}]G(z)$ represents the number of different grafts that can be realized at each node.

Let us first focus on the generating function $G(z)$:

$$G(z) = z\left(\sum_{m\geq g(n)} T_{m,n} z^m\right)^2.$$

Lemma 13 *The generating function $G(z)$ satisfies:*

$$\lim_{m\to+\infty} \frac{[z^m]G(z)}{[z^m]T(z)} \geq \frac{1}{2\sqrt{\pi g(n)}}\,\frac{1}{16n}.$$

Proof: Let $P(z)$ be the polynomial such that $G(z) = z\cdot(T(z)-P(z))^2$. Since $P(z)$ is a polynomial, the generating functions $T(z)$ and $G(z)$ have the same dominant singularity $\rho = \frac{1}{16n}$, and both singularities are of squareroot type. Therefore, using a transfer lemma, one can prove that

$$\lim_{m\to+\infty} \frac{[z^m]G(z)}{[z^m]T(z)} = \lim_{z\to\rho} \frac{G'(z)}{T'(z)} = 2\rho\left(T(\rho)-P(\rho)\right).$$

Let us find a lower bound for $T(\rho)-P(\rho)$:

$$T(\rho)-P(\rho) = \sum_{m\geq g(n)} \frac{(4n)^m}{2} C_m \rho^m.$$

Since $C_m \geq \frac{4^{m-1}}{\sqrt{\pi m^3}}$, for all $m>0$, we get $T(\rho)-P(\rho) \geq \sum_{m\geq g(n)} \frac{1}{8\sqrt{\pi m^3}}$. The function $m \mapsto m^{-3/2}$ is decreasing. Using a Riemann approximation, we get

$$T(\rho)-P(\rho) \geq \frac{1}{8\sqrt{\pi}} \int_{g(n)-1}^{\infty} \frac{du}{\sqrt{u^3}} \geq \frac{1}{4\sqrt{\pi g(n)}},$$

which concludes the proof. □

Now, let us come back to $N(z,u)$ by differentiating Equation (2) according to u and then by taking $u=1$. We get:

$$\begin{aligned} N_u(z,1) = &\sum_{p\geq 1}(2p-1)z^p K^{(p)}(A(z)+(2n-1)z) \\ &+ \sum_{p\geq 1} z^p(2zA'(z)+2(2n-1)z)K^{(p+1)}(A(z)+(2n-1)z) \\ &+ N_u(z,1)K'(N(z,1)+(2n-1)z). \end{aligned}$$

Observing that $N(z,1)=A(z)$ and isolating $N_u(z,1)$, we obtain

$$N_u(z,1) = \frac{zK'((2n-1)z+A(z)) + \sum_{p\geq 2}[(2p-1)z^p + z^{p-1}(2zA'(z)+2(2n-1)z)]K^{(p)}((2n-1)z+A(z))}{1-K'((2n-1)z+A(z))}.$$

All generating functions in the numerator of $N_u(z,1)$ have non-negative coefficients, therefore,

$$[z^m]N_u(z,1) \geq [z^m]\frac{[2z^2 + z(2zA'(z)+2(2n-1)z)]K''((2n-1)z+A(z))}{1-K'((2n-1)z+A(z))}.$$

Note that the generating function $K''((2n-1)z + A(z))$ has only non negative integer coefficients. Therefore, omitting it causes a further decrease of the coefficients of the generating function. So, after all these reductions we can conclude that

$$[z^m]N_u(z,1) \geq [z^m]\frac{4z^2A'(z)}{1-K'((2n-1)z+A(z))} =: [z^m]F(z). \tag{4}$$

Recall that:

$$A(z) = \frac{1-8nz-\sqrt{1-16nz}}{8}, \qquad A'(z) = n\left(\frac{1}{\sqrt{1-16nz}}-1\right), \quad \text{and} \quad K'(z) = \frac{1}{\sqrt{1-4z}}-1.$$

Thus,

$$F(z) = 4nz^2\left(\frac{1}{\sqrt{1-16nz}}-1\right)\ \frac{1}{2-\frac{1}{\sqrt{\frac{1}{2}-4(n-1)z+\frac{1}{2}\sqrt{1-16nz}}}}.$$

Thus, for all $r \geq 0$, substituting z by $z/16n$ yields

$$[z^r]F(z) = 2n\,(16n)^{r-2}\,[z^r]z^2\left(\frac{1}{\sqrt{1-z}}-1\right)\frac{2}{2-\frac{1}{\sqrt{\frac{1}{2}-\frac{1}{4}\left(1-\frac{1}{n}\right)z+\frac{1}{2}\sqrt{1-z}}}}. \tag{5}$$

Let us first focus on the last factor: Set

$$H(z) := \frac{2}{2-\frac{1}{\sqrt{\frac{1}{2}-\frac{1}{4}\left(1-\frac{1}{n}\right)z+\frac{1}{2}\sqrt{1-z}}}} = \frac{1}{1-\frac{1}{\sqrt{2-\left(1-\frac{1}{n}\right)z+2\sqrt{1-z}}}}$$

and observe that the shape of $H(z)$ implies that all its coefficients are non-negative and thus its dominant singularity must lie on the positive real axis. For $0 \leq z \leq 1$, the radicand $R := 2-\left(1-\frac{1}{n}\right)z+2\sqrt{1-z}$ is obviously always positive and strictly larger than 1. Thus, the denominator of $H(z)$ never vanishes and hence the dominant singularity of $H(z)$ is $z=1$, originating from the term $\sqrt{1-z}$ of R. Hence, $H(z)$ admits a Puiseux expansion

$$H(z) = c_0 + c_1\sqrt{1-z} + c_2(1-z) + c_3(1-z)^{3/2} + \ldots \tag{6}$$

around $z=1$, and all the coefficients c_k depend on n.

Lemma 14 *For all $k \geq 0$, $c_k \sim (-1)^k(2n)^{k+1}$, as n tends to infinity.*

Proof: Let $X = \sqrt{1-z}$, we have

$$\begin{aligned} H(z) &= \frac{1}{1-\frac{1}{(1+1/n)+2X+(1-1/n)X^2}} \\ &= \frac{1}{1-\frac{1}{\sqrt{1+1/n}}\sum_{k\geq 0}\binom{-1/2}{k}\left(\frac{2X+(1-1/n)X^2}{1+1/n}\right)^k} \\ &= \sum_{p\geq 0}\left(\frac{1}{1-\frac{1}{\sqrt{1+1/n}}}\right)^{p+1}\left(\sum_{k\geq 1}\binom{-1/2}{k}\frac{X^k}{(1+1/n)^k}\left(\frac{2+(1-1/n)X}{1+1/n}\right)^k\right)^p. \end{aligned}$$

Observing that

$$\frac{1}{1-\frac{1}{\sqrt{1+1/n}}} \sim 2n \quad \text{when } n \to +\infty$$

completes the proof. □

Now, let us turn to (3) and prove the desired lower bound. The following lemma immediately implies Theorem 10.

Lemma 15 *Let $\varepsilon > 0$. If $g(n) = \Omega(n^{2+\varepsilon})$, then there is a constant $C > 0$ such that*

$$\mu_n(\{f \text{ such that } L(f) \le g(n)\}) \ge C \tag{7}$$

for sufficiently large n.

Proof: First observe that

$$\frac{[z^{m-r}]G(z)}{[z^m]T(z)} = \frac{[z^{m-r}]G(z)}{[z^{m-r}]T(z)}\frac{[z^{m-r}]T(z)}{[z^m]T(z)} = \frac{[z^{m-r}]G(z)}{[z^{m-r}]T(z)}(16n)^{-r}.$$

Starting from (3), then applying Lemma 13, (4), and (5), we arrive at

$$\begin{aligned}
\mu_n(\{f \text{ such that } L(f) \le g(n)\}) &\ge \lim_{m\to+\infty} 4n \cdot \sum_{r=1}^{g(n)} [z^r]\partial_u N(z,1)\,\frac{[z^{m-r}]G(z)}{[z^m]T(z)} \\
&\ge \frac{2n}{\sqrt{\pi g(n)}} \cdot \sum_{r=1}^{g(n)} [z^r]\partial_u N(z,1)\,\frac{1}{(16n)^{r+1}} \\
&\ge \frac{1}{1024n\sqrt{\pi g(n)}} \cdot \sum_{r=0}^{g(n)-2} [z^r]\left(\frac{1}{\sqrt{1-z}} - 1\right) H(z) \\
&\sim \frac{1}{1024n\sqrt{\pi g(n)}} [z^{g(n)-2}] \frac{1}{(1-z)^{3/2}} H(z)
\end{aligned}$$

Using the Puiseux expansion (6) then the classical transfer lemma yielding $[z^m](1-z)^{-\alpha} \sim m^{\alpha-1}/\Gamma(\alpha)$, we obtain

$$[z^m]\frac{H(z)}{(1-z)^{3/2}} = c_0\frac{2\sqrt{m}}{\sqrt{\pi}} + c_2\frac{1}{\sqrt{\pi g(n)}} + O(g(n)^{-3/2}).$$

The constant hidden in the error term depends on n, but using the knowledge on the growth rate of the c_k in (6), we conclude that

$$\mu_n(\{f \text{ such that } L(f) \le g(n)\}) \ge \frac{1}{256\pi} + \frac{n^2}{128\pi g(n)} + R_n$$

with $|R_n| = O\left(n^4/g(n)^2\right)$; and here the O-constant is independent of n. □

4 Conclusion

The new specification for `And/Or` trees presented in the present paper has permitted us to count precisely valid expansions of trees of size at most n^2. We have shown by precise analytic combinatorics that the asymptotic proportion of such expanded trees is positive, implying that the `And/Or` tree distribution do not exhibit Shannon effect.

It is interesting to note that, as in the implicational systems, we need a subset of functions of quadratic complexity in order to disprove the presence of the Shannon effect.

Note that we have shown that the asymptotic probability of functions of quadratic complexity is positive, but we have not proven that it is equal to one. In view of the absence of the Shannon effect and Kozik's result on the limiting distributions it seems reasonable that asymptotically almost all functions have polynomial complexity. It appears reasonable to conjecture that the asymptotic proportion of the class we used to prove our result is one, but proving this conjecture would need further arguments.

To be more precise, it would be interesting to extend Lemma 15 in two directions: showing that it still holds if ε is 0; and proving that $C = 1$.

References

[CFGG04] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger. And/Or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5):475–497, July-September 2004.

[CGM11] B. Chauvin, D. Gardy, and C. Mailler. The growing tree distribution for Boolean functions. In *8th SIAM Workshop on Analytic and Combinatorics (ANALCO)*, pages 45–56. DMTCS proceedings, 2011.

[FGGG12] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. The fraction of large random trees representing a given boolean function in implicational logic. *Random Structures and Algorithms*, 40(3):317–349, 2012.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge U.P., Cambridge, 2009.

[GG10] A. Genitrini and B. Gittenberger. No Shannon effect on probability distributions on Boolean functions induced by Boolean expressions. In *Proc. of Analysis of Algorithms*, pages 303–316, Wien, Austria, July 2010. DMTCS Proceedings AM.

[GGKM] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler. Associative and commutative tree representations for Boolean functions. *Submitted*. `http://arxiv.org/abs/1305.0651`.

[GGKM12] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler. Probabilities of boolean functions given by random implicational formulas. *Electronic Journal of Combinatorics*, 19(2):P37, 20 pages, (electronic), 2012.

[Koz08] J. Kozik. Subcritical pattern languages for And/Or trees. In *Fifth Colloquium on Mathematics and Computer Science*, Blaubeuren, Germany, September 2008. DMTCS Proceedings.

[LS97] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas. *Random Structures and Algorithms*, 10:337–351, 1997.

[Lup62] O. B. Lupanov. Complexity of formula realization of functions of logical algebra. *Problemy Kibernetiki, 3:61-80, 1960. English translation: Problems of Cybernetics, Pergamon Press*, 3:782–811, 1962.

[PVW94] J. B. Paris, A. Vencovská, and G. M. Wilmers. A natural prior probability distribution derived from the propositional calculus. *Annals of Pure and Applied Logic*, 70:243–285, 1994.

[RS42] J. Riordan and C. E. Shannon. The number of two terminal series-parallel networks. *Journal of Math. Physics*, 21:83–93, 1942.