The Maximal Path Length of Binary Trees *

Helen Cameron† Derick Wood†

Abstract

We further refine the bounds on the path length of binary trees of a given size by considering not only the size of a binary tree, but also its height and fringe thickness (the difference between the length of a shortest root-to-leaf path and the height). We characterize the maximum path length binary trees of a given height, size, and fringe thickness. Using this characterization, we give an algorithm to find the maximum path length binary trees of a given size and fringe thickness.

1 Introduction

The path length of a tree is the sum of the lengths of the paths from the root to each node in the tree. When divided by the number of nodes in the tree, we get the average length of a path from the root to a node. Since the number of comparisons needed to find an element in a search tree is the length of the path from the root to the element's node, the average path length of a tree gives the average number of comparisons taken by an insertion, deletion, or member operation on the tree. Thus, one way we can measure the efficiency of a class of trees is by studying the path length of trees of a given size.

Knuth [Knu73] showed that a binary tree has the minimum path length among all binary trees with $N+1$ external nodes (nodes with no children) if and only if the external nodes appear on exactly two levels in the tree and those two levels are consecutive; see Figure 1. The external path length of such a tree is

$$(N+1)(\log_2(N+1)+1+\theta-2^{\theta}),$$

where $\theta = \lceil \log_2(N+1) \rceil - \log_2(N+1) \in [0,1)$. A binary tree has the maximum path length among all binary trees of size N if and only if it has

*This work was supported under a Natural Sciences and Engineering Research Council of Canada Grant No. A-5692 and under a grant from the Information Technology Research Centre.

†Data Structuring Group, Department of Computer Science, University of Waterloo, WATERLOO, Ontario N2L 3G1, CANADA

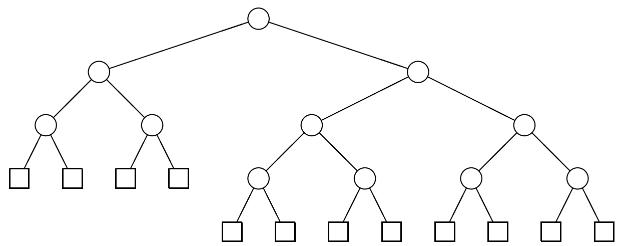

Figure 1: A minimum path length binary tree of size 11.

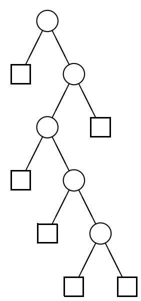

Figure 2: A maximum path length binary tree of size 5.

at most one internal node per level; that is, a binary tree has the maximum path length if and only if every internal node has at most one internal child; for example, see the tree in Figure 2. The external path length of such a tree is

$$\frac{N(N+3)}{2}.$$

The path lengths of most binary trees fall somewhere in the middle of this range, rather than at the extremes; therefore, there have been attempts to refine these bounds. Nievergelt and Wong [NW73] give an upper bound for the path length of a binary tree T in terms of the weight (the number of external nodes) and the maximum weight balance of T's subtrees. Using less information about a tree, Klein and Wood [KW89] derive the upper bound

$$(N+1)(\log_2(N+1) + \Delta - \log_2 \Delta - \Psi(\Delta))$$

for the external path length of a binary tree of size N and fringe thickness Δ, where $\Psi(\Delta) \geq 0.6622\ldots$. For a given size N and fringe thickness Δ, Klein and Wood provide a characterization of binary trees that achieve the bound up to an $O(N)$ term, if $\Delta \leq \sqrt{N+1}$.

We would like to characterize the maximum path length binary trees of a given size and fringe thickness. However, what we provide here is an algorithmic solution rather than a characterization. We characterize

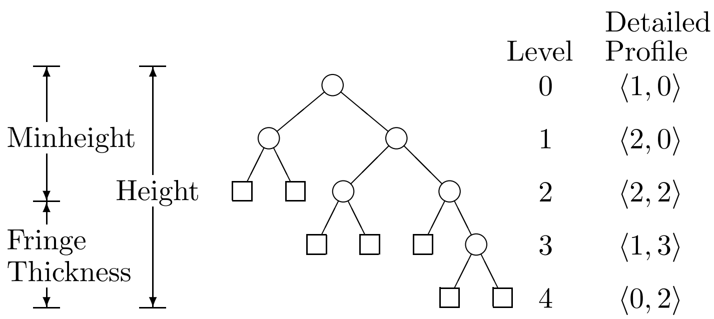

Figure 3: An example binary tree. It has size 6, height 4, minheight 2 (that is, it has a $Bin(2)$ prefix), and fringe thickness 2. Furthermore, its EPL is 21.

the maximum path length binary trees of a given height, size, and fringe thickness, and use this characterization in an algorithm that finds the height that maximizes the path length for a given size and fringe thickness. In Section 2, we define most of the terms used, introduce profiles of binary trees and prove two theorems about profiles. In Section 3, we characterize the maximum path length binary trees of a given height, size, and fringe thickness. We first introduce a simple transformation that increases the path length of a binary tree that has certain properties. Thus, a maximum path length tree of a given height, size, and fringe thickness cannot have these properties. Using these properties, we show that the numbers of external nodes on certain levels of the trees are the digits of the greedy representation of a certain number in the pseudo-binary number system, in which a number is represented as a sum of the form $\sum_{i \geq 0} a_i(2^{i+1} - 1)$, where $a_i = 0, 1$, or 2 and then we complete the characterization. In Section 4, we give an algorithm to compute the maximum path length binary trees of a given size and fringe thickness. Finally, in Section 5, we provide some open problems.

2 Definitions

We provide the basic definitions and results for binary search trees. Many of the following definitions are illustrated in Figure 3. The trees that we consider are extended trees; that is, the nodes of each tree are divided into two types: *internal nodes* (nodes that have at least one child each) and *external nodes* (nodes with no children). A *binary tree* is a tree in which every internal node has exactly two children.

The *size* of a binary tree T is the number of internal nodes in the tree; it is denoted by $size(T)$. The *height* of a tree T is the number of edges on a longest root-to-external-node path; it is denoted by $ht(T)$. The *level* of a node in a tree is the distance of the node from the root of the tree, where the distance is the number of edges on the path from the root to the node. Thus, the root is at level 0, its children (if any) are at level 1, their children are at level 2, and so on.

Definition 2.1 *The* minheight *of binary tree T, denoted by $minht(T)$, is the minimum level containing an external node; that is, $minht(T)$ is the number of edges on a shortest path from the root to an external node.*

Definition 2.2 *The* fringe thickness *of a tree T is the difference between the lengths of a longest and a shortest path from the root to an external node; that is, the fringe thickness is $ht(T) - minht(T)$.*

Note that if we are given any two of the height, minheight, and fringe thickness, then we can calculate the third value.

When we are considering the path length of a binary tree, we are interested in how far each node is from the root of the tree. We do not need to know how the nodes are arranged to achieve these distances. Using the detailed profile of a tree, we focus on the numbers of internal and external nodes on each level without worrying about the positions of the nodes.

Definition 2.3 *The* detailed profile *of binary tree T of height h is the sequence of pairs of integers $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \langle \iota_1, \epsilon_1 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$, where T has ι_i internal nodes on level i and ϵ_i external nodes on level i, for $0 \leq i \leq h$.*

But how can we prove anything about trees without taking into account the arrangement of the nodes? We can distinguish between the sequences of pairs of integers that are the detailed profiles of binary trees and the sequences that are not, as the following result demonstrates.

Theorem 2.1 *Let $\pi = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ be a sequence of integer pairs, for some $h > 0$. Then, π is the detailed profile of some non-empty binary tree T of height h if and only if*

1. *$\iota_0 = 1$ and $\epsilon_0 = 0$;*
2. *$\iota_h = 0$ and $\epsilon_h > 0$;*
3. *$\iota_i \geq 0$ and $\epsilon_i \geq 0$, for $0 \leq i \leq h$; and*
4. *$2 \cdot \iota_i = \iota_{i+1} + \epsilon_{i+1}$, for $0 \leq i < h$.*

Proof: Only If: Assume that π is the detailed profile of some non-empty binary tree of height h. Every non-empty binary tree has a root and no external nodes on level 0, so π must satisfy the first condition. A binary tree of height h has no nodes on levels greater than h, so such a tree has no internal nodes on level h because such an internal node would have to have two children on level $h+1$. Furthermore, a tree of height h must have at least two external nodes on level h. Thus, the second condition holds for π. Clearly, no binary tree can have a negative number of external or internal nodes on any level, so the third condition holds for π. Since each internal node has two children on the level beneath it, each external node has no children, and each node (except the root) must have a parent on the level above it, the fourth condition must also hold for π. Thus, if π is the detailed profile of some non-empty binary tree of height h, then π must satisfy the four conditions.

If: Assume that π is a sequence of integer pairs that satisfies the above four conditions. We show that π is the detailed profile of a non-empty binary tree of height h by constructing such a tree. We simply create the specified number of each kind of node on each level, then pair the nodes on each level (except level 0) and make each pair the two children of one of the internal nodes on the previous level. Since $\iota_i \geq 0$ and $\epsilon_i \geq 0$, for $0 \leq i \leq h$, we have a non-negative number of nodes of each kind on each level. The pairing of nodes on a level and the assigning of an internal node parent to each pair is possible because $2 \cdot \iota_i = \iota_{i+1} + \epsilon_{i+1}$, for $0 \leq i < h$. Since $\iota_0 = 1$ and $\epsilon_0 = 0$, we have a non-empty binary tree with an internal root and no other node on level 0. Because $\iota_h = 0$ and $\epsilon_h > 0$, the tree has height h. □

The perfect binary tree of height h (denoted by $Bin(h)$) is the only binary tree of height h whose external nodes all appear on one level. A recursive definition of $Bin(h)$ is given in Figure 4. It is well-known that $Bin(h)$ has size $2^h - 1$ and that this tree has the maximum number of internal nodes among all binary trees of height h. Level i of $Bin(h)$ contains 2^i nodes and each node on that level is the root of a $Bin(h-i)$ subtree.

A *snake of height* h, denoted by $Snake(h)$, is any binary tree of height h that consists of a chain of h internal nodes, one on each of the levels $0, \ldots, h-1$. See Figure 5 for an example of a $Snake(h)$. A $Snake(h)$ tree has size h, the smallest possible size for a binary tree of height h.

Definition 2.4 *A binary tree has a* binary prefix of height b*, that is, a* $Bin(b)$ *prefix, if it has* b *contiguous levels starting at the root (levels* $0, \ldots, b-1$*) that contain only internal nodes and level* b *contains at least one external node.*

Thus, the detailed profile of a binary tree with a $Bin(b)$ prefix satisfies $\iota_i = 2^i$ and $\epsilon_i = 0$, for all $0 \leq i < b$, and $\epsilon_b > 0$. Since the root of every non-empty binary tree is an internal node, every non-empty binary tree has at least a $Bin(1)$ prefix. Note that the height of the binary prefix of a binary tree is the minheight of the tree.

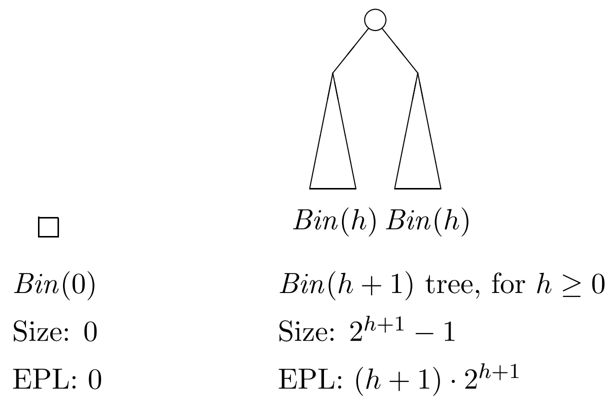

Figure 4: A recursive definition of the $Bin(h)$ tree.

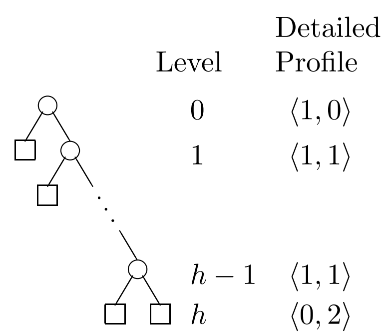

Figure 5: A snake of height h. A $Snake(h)$ has size h and its EPL is $h \cdot (h + 3)/2$.

Let T be a binary tree and let $\pi = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ be its detailed profile. The *external path length* of T is denoted by EPL(T) and is defined to be $EPL(T) = \sum_{i=0}^{h} i \cdot \epsilon_i$, In other words, the external path length is the sum, over all external nodes in the tree, of their distances from the root.

To characterize the maximum path length binary trees of a given height h, size N, and fringe thickness Δ, we determine the number of internal nodes on level $h - \Delta$ and the number of external nodes on each of the levels $h - \Delta + 1, \ldots, h - 1$. We now show that this description is equivalent to the detailed profile of the tree.

Theorem 2.2 *A binary tree T has height h, fringe thickness Δ, exactly r internal nodes on level $h - \Delta$, where $0 < r < 2^{h-\Delta}$, and e_i external nodes on level i, for $h - \Delta < i < h$, if and only if its detailed profile $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ satisfies*

- $\iota_j = 2^j$ *and* $\epsilon_j = 0$, *for all* $0 \leq j < h - \Delta$;
- $\iota_{h-\Delta} = r$ *and* $\epsilon_{h-\Delta} = 2^{h-\Delta} - r$;
- $\iota_j = 2 \cdot \iota_{j-1} - e_j$ *and* $\epsilon_j = e_j$, *for* $h - \Delta < j < h$; *and*
- $\iota_h = 0$ *and* $\epsilon_h = 2 \cdot \iota_{h-1}$.

That is, given the height h, the fringe thickness Δ, the number r of internal nodes on level $h - \Delta$, and the number e_j of external nodes on level j, for all $h - \Delta < j < h$, of a binary tree T, we can recover its detailed profile, and vice versa.

Proof: Clearly, we can compute the values of h, Δ, r, and e_i, given the detailed profile of T. Conversely, if we are given h, Δ, r, and e_i, for all $h - \Delta < i < h$, we can recover the detailed profile of T. Because T has height h and fringe thickness Δ, it must have minheight $h - \Delta$. Immediately, we must have $\iota_j = 2^j$ and $\epsilon_j = 0$, for all $0 \leq j < h - \Delta$. Since there are r internal nodes on level $h - \Delta$, we have $\iota_{h-\Delta} = r$. Because there are e_i external nodes on level i, then $\epsilon_i = e_i$, for all $h - \Delta < i < h$. We must have $\iota_h = 0$. To complete the profile, we compute $\epsilon_{h-\Delta}$, ι_i, for all $h - \Delta < i < h$, and ϵ_h as follows. Since T is a binary tree, we can use the fourth condition of Theorem 2.1: $2 \cdot \iota_i = \iota_{i+1} + \epsilon_{i+1}$, for all $0 \leq i < h$. Thus, from $\iota_{h-\Delta-1}$ and $\iota_{h-\Delta}$, we compute $\epsilon_{h-\Delta} = 2 \cdot \iota_{h-\Delta-1} - \iota_{h-\Delta}$ and from $\iota_{h-\Delta}$ and $\epsilon_{h-\Delta+1}$, we compute $\iota_{h-\Delta+1}$. By repeating this process, we can compute $\iota_{h-\Delta+2}, \iota_{h-\Delta+3}, \ldots, \iota_{h-1}$, and ϵ_h. $\square$

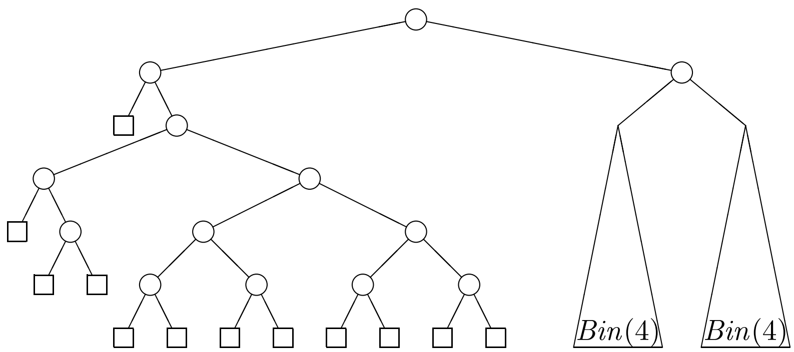

Figure 6: A maximum path length binary tree of height 6, size 43, and fringe thickness 4.

3 Maximum Path Length Binary Trees

We now characterize the maximum path length binary trees of a given height, size, and fringe thickness by giving their detailed profile. Suppose we want to describe a maximum path length binary tree of height 6, size 43, and fringe thickness 4. The tree in Figure 6 is a maximum path length binary tree of this height, size, and fringe thickness that we computed using a dynamic programming algorithm; its EPL is 256. Its nodes are placed as far from the root as possible, increasing the path length as much as possible, while maintaining the required height and fringe thickness. The idea of placing nodes as far from the root as possible enables us to find the general description of a maximum path length binary tree of a given height, size, and fringe thickness.

Definition 3.1 *Let* $MaxEPL(h, N, \Delta)$ *be the set of binary trees with the maximum path length for a given height* h*, size* N*, and fringe thickness* Δ*.*

What do such trees look like? Clearly, if $\Delta = 0$, the tree must be a $Bin(h)$ tree. Therefore, in the rest of this section, we assume that $\Delta > 0$. We deduce where external nodes may appear in a binary tree T in $MaxEPL(h, N, \Delta)$. Tree T has height h and fringe thickness Δ, so it must have a $Bin(h - \Delta)$ prefix. Therefore, there are no external nodes on the levels $0, \ldots, h - \Delta - 1$ and at least one external node on level $h - \Delta$. Furthermore, T must have at least two external nodes on level h. But what of the levels between the levels $h - \Delta$ and h?

3.1 The Distribution of External Nodes

We give two restrictions on the numbers of external nodes that can appear on the levels between $h - \Delta$ and h. We first show that there can be no more

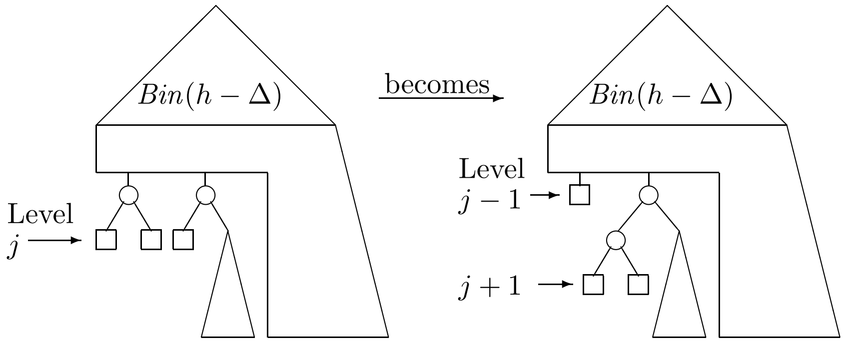

Figure 7: The transformation used in Lemma 3.1.

than two external nodes on each of these levels.

Lemma 3.1 *Let $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ be the detailed profile of a binary tree T in MaxEPL(h, N, Δ). Then, $\epsilon_i \leq 2$, for $h - \Delta < i < h$; that is, there are at most two external nodes on level i of T, for $h - \Delta < i < h$.*

Proof: We argue by contradiction. Assume that there are more than two external nodes on level j in T, for some $h - \Delta < j < h$. Then, $\epsilon_j > 2$. We show that there is a binary tree T' with detailed profile $\pi' = \langle \iota'_0, \epsilon'_0 \rangle, \ldots, \langle \iota'_h, \epsilon'_h \rangle$, where

- $\iota'_i = \iota_i$ and $\epsilon'_i = \epsilon_i$, for $0 \leq i < j - 1$ and $j + 1 < i \leq h$,
- $\iota'_{j-1} = \iota_{j-1} - 1$ and $\epsilon'_{j-1} = \epsilon_{j-1} + 1$,
- $\iota'_j = \iota_j + 1$ and $\epsilon'_j = \epsilon_j - 3$,
- $\iota'_{j+1} = \iota_{j+1}$ and $\epsilon'_{j+1} = \epsilon_{j+1} + 2$,

such that T' has the same height, size, and fringe thickness as T, but a larger EPL than T. (This corresponds to the transformation pictured in Figure 7. If no two external nodes on level j have the same parent, we exchange the positions of an external node on level j and the sibling node of another external node on level j to create a pair of external nodes on level j with the same parent before applying the transformation.) Thus, we shall have obtained a contradiction (since T is in *MaxEPL*(h, N, Δ)) and T cannot have more than two external nodes on any level between $h - \Delta$ and h.

To show that the sequence π' is the detailed profile of a binary tree, we must show that the four conditions of Theorem 2.1 hold. But most of the conditions already hold over most of π', since π' is almost the same as π, which is the detailed profile of a binary tree. We must show that the third condition still holds on the levels $j-1$, j, and $j+1$. Trivially, ϵ'_{j-1}, ι'_j, ι'_{j+1}, and ϵ'_{j+1} are nonnegative since their values are the same as or larger than the corresponding values in π. Since $\epsilon_j > 2$, we have $\epsilon'_j = \epsilon_j - 3 \geq 0$. Also, there must be at least two internal nodes on level $j-1$ in T that are the parents of the three or more external nodes on level j. Thus, $\iota'_{j-1} = \iota_{j-1} - 1 \geq 0$. We must also show that the fourth condition holds for levels $j-2$, $j-1$, and j.

Level $j-2$**:** We have $\iota'_{j-1} + \epsilon'_{j-1} = \iota_{j-1} + \epsilon_{j-1}$, from the definitions of ι'_{j-1} and ϵ'_{j-1}. Now each pair of nodes on level $j-1$ in T has an internal node parent on level $j-2$, so $\iota_{j-1} + \epsilon_{j-1} = 2 \cdot \iota_{j-2}$. Finally, using the definition of ι'_{j-2}, we have $2 \cdot \iota_{j-2} = 2 \cdot \iota'_{j-2}$. Thus, $\iota'_{j-1} + \epsilon'_{j-1} = 2 \cdot \iota'_{j-2}$, as required.

Level $j-1$**:** Using a similar argument to the one we used above for level $j-2$, we have $\iota'_j + \epsilon'_j = \iota_j + \epsilon_j - 2 = 2 \cdot \iota_{j-1} - 2 = 2 \cdot \iota'_{j-1}$.

Level j**:** In this case, we have $\iota'_{j+1} + \epsilon'_{j+1} = \iota_{j+1} + \epsilon_{j+1} + 2 = 2 \cdot \iota_j + 2 = 2 \cdot \iota'_j$.

Since the four conditions of Theorem 2.1 hold for π', it must be the detailed profile of some binary tree T'. Since $\epsilon'_h \geq \epsilon_h > 0$ and $\iota'_h = \iota_h = 0$, T and T' both have height h. Since $h - \Delta < j$, the profiles of T and T' are identical for the first $h - \Delta$ levels. Therefore, since T has fringe thickness Δ, tree T' has fringe thickness Δ, too. Finally, T' has the same size as T since it has the same number of internal nodes as T.

Finally, we compare the EPL of T and T'. Since two external nodes were moved from level j to level $j+1$ and one external node was moved from level j to level $j-1$ in the transformation from T to T', we have

$$\begin{aligned} EPL(T') &= EPL(T) - 3 \cdot j + (j-1) + 2 \cdot (j+1) \\ &> EPL(T), \end{aligned}$$

which gives a contradiction. □

Returning to the example maximum path length binary tree of height 6, size 43, and fringe thickness 4 in Figure 6, note that it agrees with the pattern predicted by Lemma 3.1; there are at most two external nodes on each of the levels 3, 4, and 5.

Second, we show that if one of the levels between $h - \Delta$ and h contains two external nodes, then no external nodes can appear on the levels beneath it, apart from level h.

Lemma 3.2 *Let $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ be the detailed profile of a binary tree T in $MaxEPL(h, N, \Delta)$. If $\epsilon_j = 2$, for some $h - \Delta < j < h$, then*

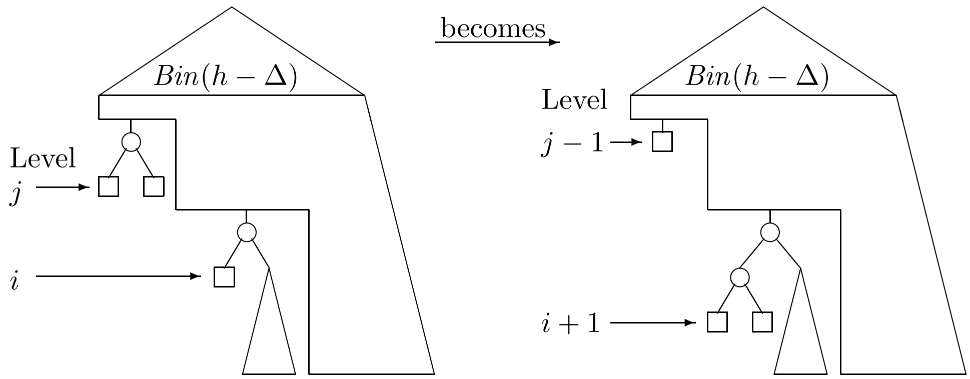

Figure 8: The transformation used in Lemma 3.2.

$\epsilon_i = 0$, for all $j < i < h$; that is, if there are two external nodes on level j, for some $h - \Delta < j < h$, in T, then there are no external nodes on the levels $j + 1, \ldots, h - 1$.

Proof: We again argue by contradiction. Assume that T has two external nodes on level j and that there is an external node on level i, for some $j < i < h$. Thus, $\epsilon_i > 0$. We show that there is a binary tree T' with the detailed profile $\pi' = \langle \iota'_0, \epsilon'_0 \rangle, \ldots, \langle \iota'_h, \epsilon'_h \rangle$, where

- $\iota'_k = \iota_k$ and $\epsilon'_k = \epsilon_k$, for all $0 \le k < j - 1$ and $j < k < i$ and $i + 1 < k \le h$;
- $\iota'_{j-1} = \iota_{j-1} - 1$ and $\epsilon'_{j-1} = \epsilon_{j-1} + 1$;
- $\iota'_j = \iota_j$ and $\epsilon'_j = \epsilon_j - 2$;
- $\iota'_i = \iota_i + 1$ and $\epsilon'_i = \epsilon_i - 1$;
- $\iota'_{i+1} = \iota_{i+1}$ and $\epsilon'_{i+1} = \epsilon_{i+1} + 2$,

such that T' has the same height, size, and fringe thickness as T, but larger EPL than T. (This corresponds to the transformation pictured in Figure 8. If the two external nodes on level j do not have the same parent, we exchange the positions of one of the external nodes on level j and the sibling node of the other external node on level j before applying the transformation.)

First, we show that π' is the detailed profile of a binary tree. The conditions of Theorem 2.1 already hold for most of π', since it is almost the same as π, which is the detailed profile of a binary tree. So, we have to show that the third and fourth conditions hold for levels $j - 1$, j, i, and $i + 1$.

Consider the third condition of Theorem 2.1. Obviously, ϵ'_{j-1}, ι'_j, ι'_i, ι'_{i+1}, and ϵ'_{i+1} are nonnegative, since their counterparts in π are nonnegative.

Since $\epsilon_j = 2$ and $\epsilon_i \geq 1$, we have $\epsilon'_j = 0$ and $\epsilon'_i \geq 0$. Furthermore, since $\epsilon_j = 2$, there must be at least one internal node on level $j-1$ that is the parent of the two external nodes. Thus, $\iota'_{j-1} = \iota_{j-1} - 1 \geq 0$.

Now consider the fourth condition of Theorem 2.1. Clearly, $2 \cdot \iota'_j = \iota'_{j+1} + \epsilon'_{j+1}$ and $2 \cdot \iota'_{i+1} = \iota'_{i+2} + \epsilon'_{i+2}$, since none of these values were changed by the transformation. On level $j-1$, we have $2 \cdot \iota'_{j-1} = 2 \cdot \iota_{j-1} - 2$, by the definition of ι'_{j-1}. The fourth condition holds for π, in particular on level $j-1$, so we have $2 \cdot \iota_{j-1} - 2 = \iota_j + \epsilon_j - 2$. By the definitions of ι'_j and ϵ'_j, we have $\iota_j + \epsilon_j - 2 = \iota'_j + \epsilon'_j$. Thus, $2 \cdot \iota'_{j-1} = \iota'_j + \epsilon'_j$; that is, the fourth condition holds for π' on level $j-1$. Similarly,

$$\begin{aligned} 2 \cdot \iota'_{i-1} &= 2 \cdot \iota_{i-1} && \text{by the definition of } \iota'_{i-1} \\ &= \iota_i + \epsilon_i - 1 + 1 && \text{since } \pi \text{ is a detailed profile} \\ &= \iota'_i + \epsilon'_i && \text{by the definitions of } \iota'_i \text{ and } \epsilon'_i. \end{aligned}$$

Also,

$$\begin{aligned} 2 \cdot \iota'_i &= 2 \cdot \iota_i + 2 && \text{by the definition of } \iota'_i \\ &= \iota_{i+1} + \epsilon_{i+1} + 2 && \text{since } \pi \text{ is a detailed profile} \\ &= \iota'_{i+1} + \epsilon'_{i+1} && \text{by the definitions of } \iota'_{i+1} \text{ and } \epsilon'_{i+1}. \end{aligned}$$

Finally,

$$\begin{aligned} 2 \cdot \iota'_{j-2} &= 2 \cdot \iota_{j-2} && \text{by the definition of } \iota'_{j-2} \\ &= \iota_{j-1} + \epsilon_{j-1} + 1 - 1 && \text{since } \pi \text{ is a detailed profile} \\ &= \iota'_{j-1} + \epsilon'_{j-1} && \text{by the definitions of } \iota'_{j-1} \text{ and } \epsilon'_{j-1}. \end{aligned}$$

Thus, the fourth condition holds for π' and, therefore, π' is the detailed profile of some binary tree T'.

Next, we show that T and T' have the same height, size, and fringe thickness. Since $h - \Delta < j < i < h$, the height and the fringe thickness are unchanged by the transformation. The numbers of internal and external nodes are not changed, although some of their positions are changed. Thus, the resulting tree T' has the same height, size, and fringe thickness as the original tree T.

Finally, we show that the EPL of T has increased. Because we moved two external nodes from level j to level $i+1$ and one external node from level i to level $j-1$, we have

$$\begin{aligned} EPL(T') &= EPL(T) - 2 \cdot j + 2 \cdot (i+1) - i + (j-1) \\ &= EPL(T) + i - j + 1 \\ &> EPL(T), \end{aligned}$$

since $j < i$. Once again we have constructed a binary tree T' with the same height, size, and fringe thickness as tree T, but with larger EPL. This is a contradiction, since T has the maximum EPL for a binary tree with its height, size, and fringe thickness. Therefore, our assumption that there is an external node on level i of tree T, for some $j < i < h$, is false. □

Corollary 3.3 *Let $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ be the detailed profile of a binary tree T in MaxEPL(h, N, Δ). If $\epsilon_j = 2$, for some $h - \Delta < j < h$, then $\epsilon_i \leq 1$, for all $h - \Delta < i < j$. That is, if there are two external nodes on level j, for some $h - \Delta < j < h$, then there is at most one external node on each of the levels $h - \Delta + 1, \ldots, j - 1$.*

Proof: By Lemma 3.1, there can be at most two external nodes on each of the levels $h-\Delta+1, \ldots, j-1$. Assume level i, for some $h-\Delta < i < j$, contains two external nodes. Then, by Lemma 3.2, levels $i+1, \ldots, h-1$ contain no external nodes. But level j contains two external nodes, a contradiction. Thus, if level j contains two external nodes, for some $h - \Delta < j < h$, then level i contains at most one external node, for all $h - \Delta < i < j$. □

Thus, external nodes are placed on the levels between (but not including) levels $h - \Delta$ and h in one of two ways. One way is that each level contains at most one external node. Alternatively, one of these levels, level j, say, contains two external nodes, those levels below level j contain no external nodes, and each level above level j contains at most one external node.

Returning again to the example maximum path length binary tree in Figure 6, note that the numbers of external nodes on each of the levels between levels $h-\Delta$ and h follow the pattern predicted by Corollary 3.3. In this case, $h = 6$ and $\Delta = 4$. Level 5 contains two external nodes; therefore, levels 3 and 4 can contain at most one external node each. In fact, there are no external nodes on level 3, and one external node on level 4.

If we consider the numbers ϵ_i of external nodes on the levels $h - \Delta + 1, \ldots, h - 1$ as the "digits" of a number in some number system, where $\epsilon_{h-\Delta+1}$ is the highest order digit and ϵ_{h-1} is the lowest order digit, we see that the digits are all either 0, 1 and 2, and if a digit 2 occurs, all of the digits in lower order positions than the 2 are 0. In fact, these sequences of ϵ_i are representations of numbers in the *pseudo-binary number system*. The pseudo-binary number system uses the digits 0, 1, and 2, and the i^{th} digit of a pseudo-binary representation is the coefficient of $2^{i+1} - 1$. (The least significant digit corresponds to index 0 and we count up from there.) For example, the decimal value of the pseudo-binary representation 201 is Value(201) $= 2 \cdot (2^3 - 1) + 0 \cdot (2^2 - 1) + 1 \cdot (2^1 - 1) = 15$. In the pseudo-binary number system, almost every nonnegative integer has more than one pseudo-binary representation; for example, both of the pseudo-binary representations 201 and 122 have decimal value 15. However, Cameron and Wood [CW91b] show that every nonnegative integer has exactly one canonical pseudo-binary representation (defined below) and that this representation is computable.

Definition 3.2 *A pseudo-binary representation* $a_n \cdots a_0$ *is* canonical *when either none of its digits are* 2 *or some digit* a_k *is* 2 *and all lower-order digits* a_i*, for all* $0 \le i < k$*, are* 0.

The canonical pseudo-binary representations match the description we need for the sequence $\epsilon_{h-\Delta+1} \cdots \epsilon_{h-1}$. From the definition, 1101 and 10200 are canonical pseudo-binary representations, but 1210 and 202 are not. Any pseudo-binary representation that has a digit 2 and some lower order digit 1 or 2 is not canonical; all other pseudo-binary representations are canonical. Cameron and Wood [CW91b] show that the representatives produced via the greedy algorithm in the pseudo-binary number system (which they call the P2 number system) are exactly the canonical pseudo-binary representations. Thus, each integer has a unique canonical pseudo-binary representation and this representation is computable.

The sequence $\epsilon_{h-\Delta+1} \cdots \epsilon_{h-1}$ (ignoring leading zeroes) is a canonical pseudo-binary representation because levels $h-\Delta+1, \ldots, h-1$ contain at most two external nodes each and if one of them contains two external nodes then the levels below it contain no external nodes. The following result gives the value of this canonical pseudo-binary representation in terms of h, N, and Δ, thereby showing how we can compute the sequence when given h, N, and Δ.

Theorem 3.4 *Let* $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ *be the detailed profile of a binary tree* T *of height* h*, size* N*, and fringe thickness* Δ*. Let* $\epsilon_i \le 2$*, for all* $h-\Delta < i < h$*, and, if* $\epsilon_j = 2$*, for some* $h-\Delta < j < h$*, then let* $\epsilon_i = 0$*, for all* $j < i < h$*. (Note that any tree in* $MaxEPL(h, N, \Delta)$ *satisfies these conditions.) Then,* $\epsilon_{h-\Delta+1} \cdots \epsilon_{h-1}$ *(ignoring leading zeroes) is the canonical pseudo-binary representation of* $2^{h-\Delta} + \iota_{h-\Delta} \cdot (2^{\Delta} - 1) - N - 1$.

Proof: Let T' be the binary tree constructed from T by replacing each of the ϵ_j external nodes on level j with a $Bin(h-j)$ subtree, for all j, $h-\Delta < j < h$. Since $ht(T) = h$ and the newly added subtrees reach to level h and no farther, T' has height h. Tree T' has fringe thickness Δ because T does and levels $0, \ldots, h-\Delta$ are unchanged. However,

$$\begin{aligned} size(T') &= size(T) + \sum_{j=h-\Delta+1}^{h-1} \epsilon_j \cdot size(Bin(h-j)) \\ &= N + \sum_{j=1}^{\Delta-1} \epsilon_{h-j} \cdot (2^j - 1). \end{aligned} \quad (1)$$

By replacing all external nodes on levels $h-\Delta+1, \ldots, h-1$ with perfect binary trees that reach all the way to level h, each non-empty subtree rooted on level $h-\Delta$ has become a $Bin(\Delta)$ tree. Therefore,

$$\begin{aligned} size(T') &= size(\mathrm{Bin}(h-\Delta)\ prefix) + \iota_{h-\Delta} \cdot size(Bin(\Delta)) \\ &= 2^{h-\Delta} - 1 + \iota_{h-\Delta} \cdot (2^{\Delta} - 1). \end{aligned} \tag{2}$$

Combining Equations 1 and 2 gives

$$\sum_{j=1}^{\Delta-1} \epsilon_{h-j}(2^j - 1) = 2^{h-\Delta} + \iota_{h-\Delta} \cdot (2^{\Delta} - 1) - N - 1.$$

□

Let us apply Theorem 3.4 to a particular example. The example maximum path length binary tree of height 6, size 43, and fringe thickness 4 in Figure 6 has 3 internal nodes on level $h-\Delta = 2$. Since $2^{h-\Delta} + \iota_{h-\Delta} \cdot (2^{\Delta} - 1) - N - 1 = 5$ and 12 is the canonical pseudo-binary number with value 5, by Theorem 3.4, this tree has no external nodes on level 3, one external node on level 4, and two external nodes on level 5. And, indeed, the maximum path length binary tree in Figure 6 agrees with this pattern.

3.2 The Number of Internal Nodes on Level $h - \Delta$

Two necessary conditions for membership in $MaxEPL(h, N, \Delta)$ are given in Lemmas 3.1 and 3.2. We now prove that they are also sufficient. We do this by examining sets of trees that have height h, fringe thickness Δ, and at most two external nodes on each of the levels between level $h-\Delta$ and level h, such that if level j contains two external nodes, for some $h-\Delta < j < h$, then level i contains no external nodes, for all $j < i < h$.

Definition 3.3 *The set $P(h, \Delta)$, where $h, \Delta \geq 0$, is the set of binary trees of height h and fringe thickness Δ that have external nodes distributed on the levels $h-\Delta+1$ to $h-1$ in one of the two following ways:*

- *Each level contains at most one external node.*
- *One of these levels, level j, contains two external nodes, levels $h-\Delta+1$ to $j-1$ each contain at most one external node, and levels $j+1$ to $h-1$ contain no external nodes.*

By Lemmas 3.1 and 3.2, every binary tree in $MaxEPL(h, N, \Delta)$ is in $P(h, \Delta)$. Consider how many non-empty subtrees rooted on level $h-\Delta$ are in such a binary tree. Since $\Delta > 0$, it must have at least one internal node on level $h-\Delta$. Since it has a $Bin(h-\Delta)$ prefix, it must have at least one external node on level $h-\Delta$; thus, it can have at most $2^{h-\Delta}-1$ internal nodes on level $h-\Delta$. Let us further subdivide the set $P(h,\Delta)$ into subsets according to the number of internal nodes rooted on level $h-\Delta$.

Definition 3.4 *Let $P(h,\Delta,r)$, where $0<r<2^{h-\Delta}$, be the set of all binary trees in the set $P(h,\Delta)$ with exactly r internal nodes on level $h-\Delta$.*

We show that r is uniquely determined by h, w, and Δ by examining the sizes of the trees in $P(h,\Delta,r)$, for all possible r. By finding the maximum and minimum sizes of binary trees in $P(h,\Delta,r)$, we show that the size of a tree in $P(h,\Delta,r)$ is strictly less than the size of any tree in $P(h,\Delta,r+1)$. Thus, if a given size N is no larger than the maximum size and no smaller than the minimum size of trees in $P(h,\Delta,r)$, for some r, then a binary tree in $\mathit{MaxEPL}(h,N,\Delta)$ must have exactly r internal nodes on level $h-\Delta$.

Lemma 3.5 *Let T be a tree in $P(h,\Delta,r)$, for some $h\geq 0$, some $\Delta\geq 0$, and some $0<r<2^{h-\Delta}$. Then, $2^{h-\Delta}+(r-1)\cdot(2^{\Delta}-1)\leq \mathit{size}(T)\leq 2^{h-\Delta}-1+r\cdot(2^{\Delta}-1)$.*

Proof: By Theorem 3.4, if the number of external nodes on level i of T is ϵ_i, then the sequence $\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2}\cdots\epsilon_{h-1}$ is the canonical pseudo-binary representation of $2^{h-\Delta}+r\cdot(2^{\Delta}-1)-\mathit{size}(T)-1$. Therefore, the larger the value of $\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2}\cdots\epsilon_{h-1}$, the smaller the size of T. Cameron and Wood [CW91b] show that if the highest order non-zero digit in a canonical pseudo-binary representation is the coefficient of $2^{n+1}-1$, then the value of the representation is at most $2\cdot(2^{n+1}-1)$. Since the highest order non-zero digit of $\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2}\cdots\epsilon_{h-1}$ can be at most $\epsilon_{h-\Delta+1}$, we have $\text{Value}(\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2}\cdots\epsilon_{h-1})\leq 2\cdot(2^{\Delta-1}-1)$. Therefore, we have $\mathit{size}(T)\geq 2^{h-\Delta}+r\cdot(2^{\Delta}-1)-2\cdot(2^{\Delta-1}-1)-1=2^{h-\Delta}+(r-1)\cdot(2^{\Delta}-1)$.

Consider a binary tree T' that has height h, fringe thickness Δ, exactly r $\mathit{Bin}(\Delta)$ subtrees rooted on level $h-\Delta$ and no other non-empty subtrees rooted on level $h-\Delta$. Tree T' is pictured in Figure 9. Since T' has no external nodes on the levels $h-\Delta+1,\ldots,h-1$, tree T' is in $P(h,\Delta,r)$. As we showed in the proof of Theorem 3.4, any tree T in $P(h,\Delta,r)$ can be transformed into a tree with the same description as T' (the positions of the r $\mathit{Bin}(\Delta)$ subtrees may be different) by replacing each external node on level j with a $\mathit{Bin}(h-j)$ subtree, for $h-\Delta<j<h$. Thus, for any tree T in $P(h,\Delta,r)$, we have $\mathit{size}(T)\leq\mathit{size}(T')=2^{h-\Delta}-1+r\cdot(2^{\Delta}-1)$. □

In fact, the lower bound of Lemma 3.5 is tight because a binary tree of height h and fringe thickness Δ with $r-1$ $\mathit{Bin}(\Delta)$ subtrees, one $\mathit{Bin}(1)$ subtree, and no other non-empty subtrees rooted on level $h-\Delta$ is in $P(h,\Delta,r)$ and has size $2^{h-\Delta}+(r-1)\cdot(2^{\Delta}-1)$; see Figure 10.

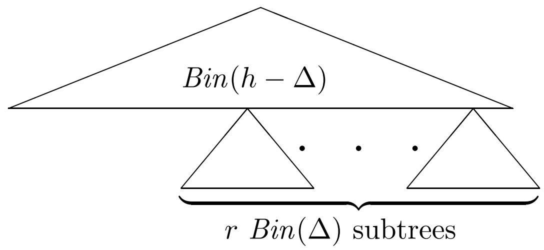

Figure 9: A maximum size tree in $P(h, \Delta, r)$.

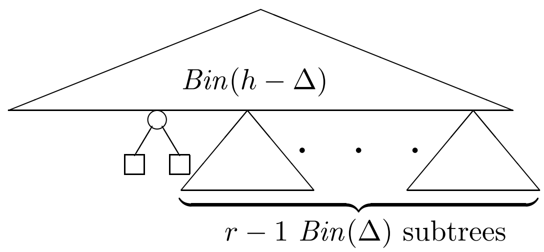

Figure 10: A minimum size tree in $P(h, \Delta, r)$.

Corollary 3.6 *Let T_r be a binary tree in $P(h, \Delta, r)$ and T_{r+1} be a binary tree in $P(h, \Delta, r+1)$. Then, $size(T_r) < size(T_{r+1})$.*

Proof: Applying Lemma 3.5, we have $size(T_r) \le 2^{h-\Delta} - 1 + r \cdot (2^{\Delta} - 1)$ and $2^{h-\Delta} + r \cdot (2^{\Delta} - 1) \le size(T_{r+1})$. Therefore, $size(T_r) < size(T_{r+1})$. □

Since the sizes of trees in $P(h, \Delta, r)$ fall within a range that does not overlap with the range of sizes in $P(h, \Delta, r')$, where $r' \neq r$, we can compute the number r of internal nodes on level $h - \Delta$ of a tree in $MaxEPL(h, N, \Delta)$ by computing the range that contains N. In fact, there is a simple formula to compute r given h, N, and Δ, as the following result shows.

Theorem 3.7 *If T is in $MaxEPL(h, N, \Delta)$, then T has exactly r internal nodes on level $h - \Delta$, where*

$$r = \left\lfloor \frac{N - 2^{h-\Delta}}{2^{\Delta} - 1} \right\rfloor + 1.$$

Proof: If T is a maximum path length binary tree of height h, size N, and fringe thickness Δ, then T is in $P(h, \Delta, r)$, for some $0 < r < 2^{h-\Delta} - 1$. Therefore, by Lemma 3.5,

$$2^{h-\Delta} + (r-1) \cdot (2^{\Delta} - 1) \le size(T) \le 2^{h-\Delta} - 1 + r \cdot (2^{\Delta} - 1).$$

We can rewrite this as

$$r - 1 \leq \frac{N - 2^{h-\Delta}}{2^{\Delta} - 1} \leq r - \frac{1}{2^{\Delta} - 1}.$$

Thus, since r is an integer and $\Delta \geq 1$, we have

$$r = \left\lfloor \frac{N - 2^{h-\Delta}}{2^{\Delta} - 1} \right\rfloor + 1.$$

Furthermore, there cannot be any other binary tree T' in $MaxEPL(h, N, \Delta)$ with a number r' of internal nodes on level $h - \Delta$ different from r. If there were such a T', then, by Corollary 3.6, either $size(T) < size(T')$ (if $r < r'$) or $size(T') < size(T)$ (if $r' < r$). But $size(T) = size(T') = N$. Thus, there can be only one choice for the number of internal nodes on level $h - \Delta$. □

Let us return to the previous example of a maximum path length binary tree of height 6, size 43, and fringe thickness 4. By Theorem 3.7, to find the number of internal nodes on level 2 of such a tree, we must find the integer r such that $2^{h-\Delta} + (r-1) \cdot (2^{\Delta} - 1) \leq 43 \leq 2^{h-\Delta} - 1 + r \cdot (2^{\Delta} - 1)$; that is, $r = \lfloor (N - 2^{h-\Delta})/(2^{\Delta} - 1) \rfloor + 1$. With $r = \lfloor 39/15 \rfloor + 1 = 3$, we have $34 \leq 43 \leq 48$. Thus, the maximum path length binary tree of height 6, size 43, and fringe thickness 4 in Figure 6 has exactly three internal nodes on level 2.

3.3 The Description of a Tree in $MaxEPL(h, N, \Delta)$

Let us summarize what we have discovered about binary trees in $MaxEPL(h, N, \Delta)$ by giving their detailed profiles.

Theorem 3.8 *Let $\pi(T) = \langle \iota_0, \epsilon_0 \rangle, \ldots, \langle \iota_h, \epsilon_h \rangle$ be the detailed profile of a binary tree T in $MaxEPL(h, N, \Delta)$. Then,*

- $\iota_i = 2^i$ *and* $\epsilon_i = 0$, *for all* $0 \leq i < h - \Delta$*;*
- $\iota_{h-\Delta} = r$ *and* $\epsilon_{h-\Delta} = 2^{h-\Delta} - r$, *where* $r = \lfloor (N - 2^{h-\Delta})/(2^{\Delta} - 1) \rfloor + 1$*;*
- $\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2} \cdots \epsilon_{h-1}$ *(ignoring leading zeros) is the canonical pseudo-binary representation of* $2^{h-\Delta} + r \cdot (2^{\Delta} - 1) - N - 1$*;*
- ι_i, *for all* $h - \Delta < i < h$, *can be found by using* $2 \cdot \iota_{i-1} = \iota_i + \epsilon_i$, *once* $\iota_{h-\Delta}$ *and* ϵ_i *are known; and*
- $\iota_h = 0$ *and* $\epsilon_h = N + 1 - (2^{h-\Delta} - r) - \sum_{i=1}^{\Delta-1} \epsilon_{h-\Delta+i}$.

Proof: The first condition follows immediately since T has height h and fringe thickness Δ. The second condition, that T has r internal nodes on level $h-\Delta$, follows directly from Theorem 3.7. The third condition follows from Theorem 3.4. Since these first three items give the number of external nodes on the levels $0,\ldots,h-1$, any remaining external nodes appear on level h. Thus, $\epsilon_h = (N+1) - (2^{h-\Delta} - r) - \sum_{i=1}^{\Delta-1} \epsilon_{h-\Delta+i}$. Of course, since T has height h, $\iota_h = 0$. Finally, the fourth condition follows from the fourth condition of Theorem 2.1. □

Returning to the previous example, suppose we are asked to characterize a maximum path length binary tree of height $h=6$, size $N=43$, and fringe thickness $\Delta = 4$. We previously found that such a tree must have three non-empty subtrees rooted on level 2, and that there are no external nodes on level 3, exactly one external node on level 4, and exactly two external nodes on level 5 of such a tree. By Theorem 3.8, we can find ι_j, for all $2 < j < 6$, using the formula $\iota_j = 2 \cdot \iota_{j-1} - \epsilon_j$. Thus, there are $2 \cdot 3 - 0 = 6$ internal nodes on level 3, $2 \cdot 6 - 1 = 11$ internal nodes on level 4, and $2 \cdot 11 - 2 = 20$ internal nodes on level 5. Also, there are $44 - (2^2 - 3) - 1 - 2 = 40$ external nodes on level 6. Thus, the detailed profile of the maximum path length binary tree is $\langle 1,0\rangle\langle 2,0\rangle\langle 3,1\rangle\langle 6,0\rangle\langle 11,1\rangle\langle 20,2\rangle\langle 0,40\rangle$. If we put all external nodes on levels 3, 4, and 5 in one of the subtrees rooted on level 2, we obtain a binary tree such as the one in Figure 6.

Theorem 3.9 *Let T be a binary tree of height h, size N, and fringe thickness Δ. Then,*

$$\mathrm{EPL}(T) \leq (N+1)\cdot h - \Delta \cdot (2^{h-\Delta} - r) - \sum_{i=1}^{\Delta-1} (\Delta - i)\cdot \epsilon_{h-\Delta+i},$$

where

- $r = \lfloor (N - 2^{h-\Delta})/(2^{\Delta} - 1) \rfloor + 1$, *and*
- $\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2}\cdots\epsilon_{h-1}$ *(ignoring leading zeros) is the canonical pseudo-binary representation of* $2^{h-\Delta} + r\cdot(2^{\Delta}-1) - N - 1$.

Proof: By Theorem 3.8, a binary tree T' in $\mathit{MaxEPL}(h,N,\Delta)$ has $2^{h-\Delta} - r$ external nodes on level $h-\Delta$, where $r = \lfloor (N - 2^{h-\Delta})/(2^{\Delta}-1)\rfloor + 1$, since it has height h, fringe thickness Δ, and r internal nodes on level $h-\Delta$. It also has $\epsilon_{h-\Delta+i}$ external nodes on level $h-\Delta+i$, for $0 < i < \Delta$, where $\epsilon_{h-\Delta+1}\epsilon_{h-\Delta+2}\cdots\epsilon_{h-1}$ is the canonical pseudo-binary representation of $2^{h-\Delta} + r\cdot(2^{\Delta}-1) - N - 1$. The remaining $N+1-(2^{h-\Delta}-r) - \sum_{i=1}^{\Delta-1}\epsilon_{h-\Delta+i}$ external nodes appear on level h. Since $\mathit{EPL}(T') = \sum_{i=0}^{h} i\cdot\epsilon_i$, the EPL of the binary tree T' is given by

$$\begin{aligned} EPL(T') &= (h-\Delta)\cdot(2^{h-\Delta}-r) + \sum_{i=1}^{\Delta-1}(h-\Delta+i)\cdot\epsilon_{h-\Delta+i} \\ &\quad + h\cdot(N+1-(2^{h-\Delta}-r) - \sum_{i=1}^{\Delta-1}\epsilon_{h-\Delta+i}) \\ &= (N+1)\cdot h - \Delta\cdot(2^{h-\Delta}-r) - \sum_{i=1}^{\Delta-1}(\Delta-i)\cdot\epsilon_{h-\Delta+i}. \end{aligned}$$

Finally, the EPL of any binary tree T of height h, size N, and fringe thickness Δ is at most as large as the EPL of a binary tree T' in $MaxEPL(h, N, \Delta)$. □

For example, consider a binary tree T of height 6, size 43, and fringe thickness 4. A maximum path length binary tree T' of height 6, size 43, and fringe thickness 4 has, as we have seen, three internal nodes on level 2, no external nodes on level 3, one external node on level 4, and two external nodes on level 5. Thus, by the above theorem, the external path length of T is bounded from above by

$$\begin{aligned} EPL(T) &\leq EPL(T') \\ &= 44\cdot(2+4) - 4\cdot(2^2-3) - ((4-1)\cdot 0 \\ &\quad + (4-2)\cdot 1 + (4-3)\cdot 2) \\ &= 256. \end{aligned}$$

4 Path Length, Size, and Fringe Thickness

Now that we have characterized the maximum path length binary trees of a given height, size, and fringe thickness, we use this characterization to compute the height that gives the maximum path length. Because the numbers of external nodes on some levels of the maximum path length binary tree of a given height, size, and fringe thickness are given by the pseudo-binary representation of a number that is a function of the height as well as the size and fringe thickness, computing the height that gives the maximum path length is not immediate. Although we cannot give a formula for the height that gives the maximum path length, we can compute it. Given a size and a fringe thickness, we compute the maximum path length for each height using Theorem 3.9 and then choose the height that gives the largest path length. We examine only the heights for which there exist binary trees of the given size and fringe thickness.

Given a size N and a fringe thickness Δ, what is the minimum height that a binary tree can have and what is the maximum height? For $\Delta = 0$ and $\Delta = 1$, there is exactly one choice of height. If $\Delta = 0$, the size must be $N = 2^h - 1$, for some $h \geq 0$, and the tree is a $Bin(h)$ tree. If $\Delta = 1$, then the height of the tree must be $\lceil \log_2(N+1) \rceil$, and level $\lfloor \log_2(N+1) \rfloor$ contains $(N+1) - 2^{\lfloor \log_2(N+1) \rfloor}$ internal nodes. Assume in what follows that $\Delta > 1$ and that *size* $> \Delta$ (we must have enough nodes for a $Snake(\Delta + 1)$ tree to obtain fringe thickness Δ).

We characterize the combinations of height, size, and fringe thickness for which binary trees exist.

Theorem 4.1 *Let h and Δ be nonnegative integers such that $\Delta < h$ and let N be a nonnegative integer. Then, there is a binary tree of height h, size N, and fringe thickness Δ if and only if*

$$2^{h-\Delta} + \Delta - 1 \leq N \leq (2^{h-\Delta} - 1) \cdot 2^{\Delta}.$$

Proof: **Only if**: Assume that there is a binary tree T of height h, size N, and fringe thickness Δ. Clearly, a binary tree of height h and fringe thickness Δ must have a $Bin(h - \Delta)$ prefix and at least one path leading from level $h-\Delta$ to level h. Thus, a minimum size binary tree of height h and fringe thickness Δ consists entirely of a $Bin(h - \Delta)$ prefix and a $Snake(\Delta)$ rooted on level $h-\Delta$; see Figure 11. Since the size of T must be at least the minimum size for a binary tree of height h and fringe thickness Δ, we have $2^{h-\Delta} + \Delta - 1 \leq N$. Consider a maximum size tree of height h and fringe thickness Δ. It must have at least one external node on level $h - \Delta$ and, therefore, no more than $2^{h-\Delta} - 1$ internal nodes on level $h - \Delta$. We show, by contradiction, that the maximum size tree must have exactly $2^{h-\Delta} - 1$ internal nodes on level $h - \Delta$. If the maximum size tree has fewer than $2^{h-\Delta} - 1$ internal nodes on level $h - \Delta$, we can replace one of the external nodes on level $h-\Delta$ with a $Bin(\Delta)$ subtree, thereby creating another binary tree of larger size than the maximum size tree, but with the same height and fringe thickness. Thus, the maximum size tree of height h and fringe thickness Δ has exactly $2^{h-\Delta} - 1$ internal nodes on level $h - \Delta$. We show that each of these internal nodes must be the root of a $Bin(\Delta)$ subtree in a similar manner. Since the height of the maximum size tree is h, if one of these internal nodes is not the root a $Bin(\Delta)$ subtree, it must be the root of some other subtree of height at most Δ. Then, we can replace the subtree by a $Bin(\Delta)$ subtree (which has the largest size among all binary trees of height at most Δ), creating another binary tree of height h and fringe thickness Δ with larger size than the maximum size tree. Thus, each of the $2^{h-\Delta} - 1$ internal nodes on level $h - \Delta$ of a maximum size tree of height h and fringe thickness Δ is the root of a $Bin(\Delta)$ subtree. An example of a maximum

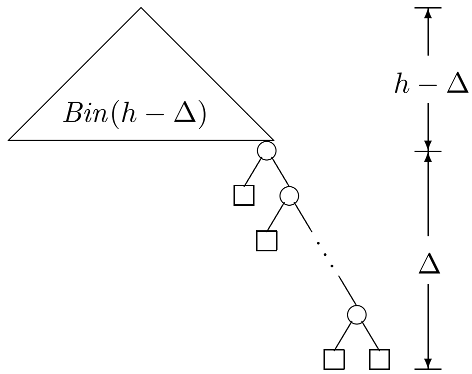

Figure 11: A minimum size tree of height h and fringe thickness Δ.

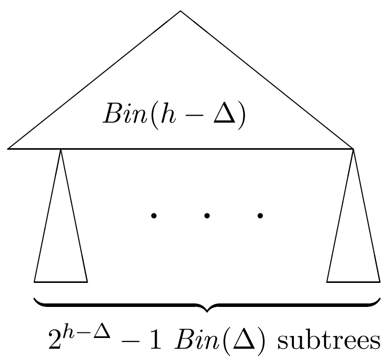

Figure 12: A maximum size tree of height h and fringe thickness Δ.

size tree of height h and fringe thickness Δ is given in Figure 12. Since the size of T can be at most the maximum size for a binary tree of height h and fringe thickness Δ, we have $2^{h-\Delta}+\Delta-1 \le N \le (2^{h-\Delta}-1)\cdot 2^{\Delta}$.

If: Assume that $2^{h-\Delta}+\Delta-1 \le N \le (2^{h-\Delta}-1)\cdot 2^{\Delta}$. We construct a binary tree of height h, size N, and fringe thickness Δ by starting with a maximum size binary tree T of height h and fringe thickness Δ and then removing nodes until we have size N, being careful to maintain the correct height and fringe thickness. Since $2^{h-\Delta}+\Delta-1 \le N$, we will have enough nodes remaining, after we have pruned the tree, to have a $Bin(h-\Delta)$ prefix and at least one path leading from level $h-\Delta$ to level h, that is, for the tree to have height h and fringe thickness Δ. To decide which nodes can be removed and which must not be removed, we choose a path P in the maximum size tree T leading from level $h-\Delta$ to level h. We mark the internal nodes on P and all nodes in the $Bin(h-\Delta)$ prefix (that is, all nodes on levels $0,\ldots,h-\Delta-1$) as "unremovable" (ensuring that we always have height h and fringe thickness Δ) and we mark all other internal nodes as "removable." Then, we perform the following deletion $(2^{h-\Delta}-1)\cdot 2^{\Delta}-N$ times: find a removable node with two external children and replace it by an external node. Clearly, if we can find such a node and replace it by an external node, then the resulting tree is a binary tree of height h and fringe thickness Δ and the size is reduced by one. We now show that if we have a binary tree of height h and fringe thickness Δ with size larger than $2^{h-\Delta}+\Delta-1$ whose nodes are marked in the manner described above, then we can find a removable node with two external children. Since all removable nodes are on levels $h-\Delta,\ldots,h-1$ and the only unremovable nodes on levels $h-\Delta,\ldots,h-1$ are on path P, all internal descendants of removable nodes are removable. Thus, if we choose any removable node v, then the last level of internal nodes of the subtree rooted at v consists of removable nodes with two external children. Thus, we can perform the deletion operation until we have pruned the maximum size tree down to size N while maintaining height h and fringe thickness Δ. That is, given integers $h \geq 0$, $N > 0$, and $\Delta \geq 0$ such that $2^{h-\Delta}+\Delta-1 \leq N \leq (2^{h-\Delta}-1)\cdot 2^{\Delta}$, there is a binary tree of height h, size N, and fringe thickness Δ. □

We can use the relation $2^{h-\Delta}+\Delta-1 \leq N \leq (2^{h-\Delta}-1)\cdot 2^{\Delta}$ to compute bounds on the height. The lower bound $\lceil \log_2(N+2^{\Delta})\rceil \leq h$ follows from $N \leq (2^{h-\Delta}-1)\cdot 2^{\Delta}$ and the upper bound $h \leq \Delta + \lfloor \log_2(N-\Delta+1)\rfloor$ follows from $2^{h-\Delta}+\Delta-1 \leq N$.

5 Concluding Remarks

We characterized the maximum path length binary trees of a given height, size, and fringe thickness. In Klein and Wood [KW89], the two binary trees shown in Figures 13 and 14 are used to show that the upper bound on the EPL of a binary tree of size N and fringe thickness Δ is tight if $\Delta < \sqrt{N+1}$ and quite sharp if $\sqrt{N+1} \leq \Delta$. Using the characterization, we see that these two binary trees have the maximum path length for their heights, sizes, and fringe thicknesses since they have at most one external node on each of the levels between the minheight and the height of the trees. However, it is not known whether these trees have the maximum path for their sizes and fringe thicknesses. The characterization of the binary trees with maximum path length for a given size and fringe thickness remains open. An area for further research is to extend these results to multiway trees. Another problem is to characterize the minimum path length binary trees for a given size N and fringe thickness Δ. De Santis and Persiano [DP91] have characterized the minimum path length binary trees whose fringe thicknesses satisfy $\Delta \leq (N+1)/2$. In [CW91a], Cameron and Wood show that, given a height as well as a size and fringe thickness without any

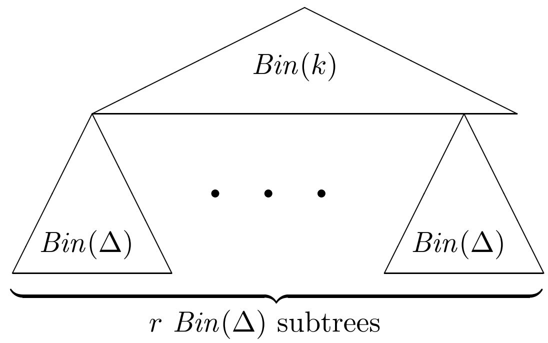

Figure 13: The tree $T_1(r, k, \Delta)$ of Klein and Wood.

restrictions, the minimum path length binary trees can be characterized.

References

[CW91a] Helen Cameron and Derick Wood. Binary trees, fringes, and minimum path lengths. Unpublished manuscript, 1991.

[CW91b] Helen Cameron and Derick Wood. Pm numbers, ambiguity, and regularity. Technical Report CS-91-47, University of Waterloo, 1991.

[DP91] Alfredo De Santis and Guiseppe Persiano. Tight upper and lower bounds on the path length of binary trees. Unpublished manuscript, 1991.

[Knu73] Donald E. Knuth. *The Art of Computer Programming, Vol. 3: Sorting and Searching.* Addison-Wesley, Reading, MA, 1973.

[KW89] Rolf Klein and Derick Wood. On the path length of binary trees. *Journal of the Association for Computing Machinery*, 36(2):280–289, 1989.

[NW73] J. Nievergelt and C. K. Wong. Upper bounds for the total path length of binary trees. *Journal of the ACM*, 20(1):1–6, January 1973.

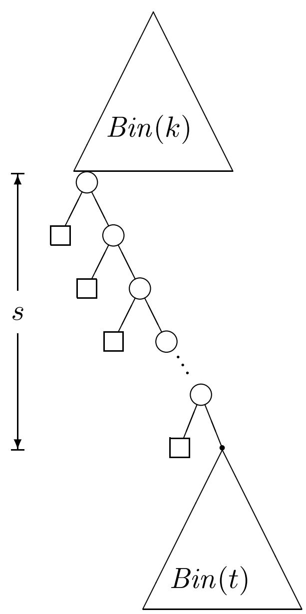

Figure 14: The tree $T_2(k, s, t)$ of Klein and Wood.