A new version of Toom’s proof

Peter Gács*
Boston University
April 20, 1995

Abstract

There are several proofs now for the stability of Toom’s example of a two-dimensional stable cellular automaton and its application to fault-tolerant computation. Simon and Berman simplified and strengthened Toom’s original proof: the present report is simplified exposition of their proof.

1 Introduction

For a 2-dimensional cellular automaton, the set \( C \) of sites is the set \( \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \) where \( \mathbb{Z}_m \) is the set of integers modulo the potentially infinite modulus \( m \). Cellular automata in other dimensions are defined similarly. In a space-time vector \((x, t)\), we will always write the space coordinate first. A configuration is a function \( \xi(x) \) assigning a state to each site from a finite set \( S \) of states. An evolution is a partial function \( \eta(x, t) \) assigning a state to each site at each time within the interval of interest. In discrete-time cellular automata, the only kind discussed here, state transitions occur only at integer times.

A (one-dimensional) deterministic cellular automaton is determined by a local transition rule \( \text{Trans}() \); we can denote it by

\[ \text{CA}(\text{Trans}). \]

An evolution \( \eta \) in one dimension is a trajectory if

\[ \eta(x, t) = \text{Trans}(\eta(x - B, t - T), \eta(x, t - T), \eta(x + B, t - T)) \]

holds for all \( x, t \). Cellular automata in several dimensions are defined similarly. Given a configuration \( \xi \) over the space \( C \) and a transition function, there is a unique trajectory \( \eta \) with the given transition function and the initial configuration \( \eta(\cdot, 0) = \xi \).

*Partially supported by NSF grant CCR-9204284
A random evolution is a pair \((\mu, \eta)\) where \(\mu\) is a measure over some measurable space \(\Omega\) and \(\eta(\omega)\) is a measurable function from \(\Omega\) to the set of evolutions. Generally, we speak about a fixed random evolution \(\eta\) and refer to \(\mu\) as Prob.

We will say that a random evolution \((\mu, \eta)\) is a trajectory of the \(\varepsilon\)-perturbation \(CA_\varepsilon(\text{Trans})\) of the transition function \(\text{Trans}\) if the following holds. For all \(x, t, r_{-1}, r_0, r_1\), if \(\eta(x + j, t - 1) = r_j\) \((j = -1, 0, 1)\) and \(\eta(x', t')\) is otherwise fixed arbitrarily for all \(t' < t\) and for all \(x' \neq x, t' = t\), then the conditional probability of \(\eta(x, t) = \text{Trans}(r_{-1}, r_0, r_1)\) is at least \(1 - \varepsilon\).

A simple stable two-dimensional deterministic cellular automaton given by Toom in [3] can be defined as follows. First we define the neighborhood

\[ H = \{(0, 0), (0, 1), (1, 0)\}. \]

The transition function is, for each cell \(x\), a majority vote over the three values \(x + g_i\) where \(g_i \in H\).

As in [2], let us be given an arbitrary one-dimensional transition function \(\text{Trans}\) and the integers \(N, T\). We define the three-dimensional transition function \(\text{Trans}'\) as follows. The interaction neighborhood is \(H \times \{-1, 0, 1\}\) with the neighborhood \(H\) defined above. The rule \(\text{Trans}'\) says: in order to obtain your state at time \(t + 1\), first apply majority voting among self and the northern and eastern neighbors in each plane defined by fixing the third coordinate. Then, apply rule \(\text{Trans}\) on each line obtained by fixing the first and second coordinates.

For a finite or infinite \(m\), let \(C\) be our 3-dimensional space that is the product of \(\mathbb{Z}_m^2\) and a 1-dimensional (finite or infinite) space \(A\) with \(N = |A|\). For a trajectory \(\zeta\) of \(\text{Trans}\) on \(A\), we define the trajectory \(\zeta'\) of \(\text{Trans}'\) on \(C\) by \(\zeta'(i, j, n, t) = \zeta(n, t)\).

Let \(\zeta'\) be a trajectory of \(\text{Trans}'\) and \(\eta\) a trajectory of \(CA_\varepsilon(\text{Trans}')\) such that \(\eta(0, w) = \zeta'(0, w)\).

(1.1) Theorem

Suppose \(\varepsilon < \frac{1}{32m^{1/2}}\). If \(m = \infty\) then

\[ \text{Prob}\left\{ \eta(w, t) \neq \zeta'(w, t) \right\} \leq 24\varepsilon. \]

If \(m\) is finite then

\[ \text{Prob}\left\{ \eta(w, t) \neq \zeta'(w, t) \right\} \leq 24m^2 N (2 \cdot (12)^{3/2} \cdot \varepsilon^{1/12})^m + 24\varepsilon. \]

The proof we give here is a further simplification of the simplified proof of [1].

Let \(\text{Noise}\) be the set of space-time points \(v\) where \(\eta\) does not obey the transition rule of \(\text{Trans}'\). Let us define a new process \(\xi\) such that \(\xi(w, t) = 0\) if \(\eta(w, t) = \zeta'(w, t)\), and 1 otherwise. Let

\[ \text{Cerr}(a, b, u, t) = \text{Maj}(\xi(a, b, u, t), \xi(a + 1, b, u, t), \xi(a, b + 1, u, t)). \]
Then for all points \((a, b, u, t + 1) \in \text{Noise}(\eta)\), we have
\[
\xi(a, b, u, t + 1) \leq \max(\text{Corr}(a, b, u - 1, t), \text{Corr}(a, b, u, t), \text{Corr}(a, b, u + 1, t)).
\]

Now, Theorem 1.1 can be restated as follows:
Suppose \(\varepsilon < \frac{1}{3212}\). If \(m = \infty\) then
\[
\text{Prob}\left\{ \xi(w, t) = 1 \right\} \leq 24\varepsilon.
\]
If \(m\) is finite then
\[
\text{Prob}\left\{ \xi(w, t) = 1 \right\} \leq 24tm^2N(2 \cdot (12)^{1/12})^m + 24\varepsilon.
\]

2 Proof using small explanation trees
If \(m < \infty\) let \(C' = \mathbb{Z}^3\) be our covering space, and \(V' = C' \times \mathbb{Z}\) our covering space-time. There is a projection \(\text{proj}(u)\) from \(C'\) to \(C\) defined by
\[
\text{proj}(u)_i = u_i \mod m \quad (i = 1, 2).
\]
This rule can be extended to \(C'\) identically. We define a random process \(\xi'\) over \(C'\) by
\[
\xi'(w, t) = \xi(\text{proj}(w), t).
\]
The set \(\text{Noise}\) is extended similarly to \(\text{Noise}'\). Now, if \(\text{proj}(w_1) = \text{proj}(w_2)\) then \(\xi'(w_1, t) = \xi'(w_2, t)\) and therefore the failures at time \(t\) in \(w_1\) and \(w_2\) are not independent.

In figures, we generally draw space-time with the time direction going down. Therefore, for two neighbors \(u, u'\) of the space \(\mathbb{Z}\) and integers \(a, b, t\), we will call arrows, or vertical edges the following kind of (undirected) edges:
\[
\{(a, b, u, t), (a, b, u', t - 1)\}, \{(a, b, u, t), (a + 1, b, u', t - 1)\},
\{(a, b, u, t), (a, b + 1, u', t - 1)\}.
\]
We will call forks, or horizontal edges the following kinds of edges:
\[
\{(a, b, u, t), (a + 1, b, u, t)\}, \{(a, b, u, t), (a, b + 1, u, t)\},
\{(a + 1, b, u, t), (a, b + 1, u, t)\}.
\]
We define the graph \(G\) by introducing all possible arrows and forks. Thus, a point is adjacent to 6 possible forks and 6 possible arrows: the degree of \(G\) is at most
\[
r = 12.
\]
(If the space is \(d + 2\)-dimensional instead of 3, then \(r = 6(d + 1)\).) We use the notation \(\text{Time}(w, t) = t\).
(2.1) **Explanation Tree Lemma**  Let $u$ be a point outside $\text{Noise'}$ with $\xi'(u) = 1$. Then there is a subtree $\text{Expl}(u, \xi')$ of $G$ rooted at $u$ called an explanation of $u$ such that if $n$ nodes of $\text{Expl}$ belong to $\text{Noise'}$ then the number of edges of $\text{Expl}$ is at most $4n - 4$.

This lemma will be proved in the next section. To use it in the proof of the main theorem, we need some easy lemmas on trees whose nodes have weights 0 or 1, with the root having weight 0. The redundancy of such a tree is the ratio of its number of edges to its weight. The set of nodes of weight 1 of a tree $T$ will be denoted by $F(T)$. A subtree of a tree is a subgraph that is a tree.

(2.2) **Lemma**  Let $T$ be a weighted tree of total weight $w > 3$ and redundancy $\lambda$. Then there is a subtree of total weight $w_1$ with $w/3 < w_1 \leq 2w/3$, and redundancy $\leq \lambda$.

**Proof:** Let us order $T$ from the root $r$ down. Let $T_1$ be a minimal subtree below $r$ with weight $w_1 = w/3$. Then the subtrees immediately below $T_1$ all weigh $\leq w/3$. Let us delete as many of these as possible while keeping $T_1$ weigh $w/3$. At this point, the weight $w_1$ of $T_1$ is $\geq w/3$ but $\leq 2w/3$ since we could subtract a number $\leq w/3$ from it so that $w_1$ would become $\leq w/3$ (note that since $w > 3$) the tree $T_1$ is not a single node.

Now $T$ has been separated by a node into $T_1$ and $T_2$, with weights $w_1, w_2 > w/3$. Since the root of a tree has weight 0 by definition the possible weight of the root of $T_1$ stays in $T_2$ and we have $w_1 + w_2 = w$. The redundancy of $T$ is then a weighted average of the redundancies of $T_1$ and $T_2$, and we can choose the one of the two with the smaller redundancy: its redundancy is smaller than that of $T$. $\square$

(2.3) **Tree Separator Theorem**  Let $T$ be a weighted tree with weight $w$ and redundancy $\lambda$, and let $k < w$. Then $T$ has a subtree with weight $w'$ such that $k/3 < w' \leq k$ and redundancy $\leq \lambda$.

**Proof:** Let us perform the operation of Lemma 2.2 repeatedly, until we get weight $\leq k$. Then the weight $w'$ of the resulting tree is $\geq k/3$. $\square$

(2.4) **Tree Counting Lemma**  In a graph of maximum node degree $r$ the number of weighted subtrees rooted at a given node and having $k$ edges is at most $2r \cdot (2r^2)^k$.

**Proof:** Let us number the nodes of the graph arbitrarily. Each tree of $k$ edges can now be traversed in a breadth-first manner. At each non-root node of the tree of degree $i$ from which we continue, we make a choice out of $r$ for $i$ and then a choice out of $r - 1$ for each of the $i - 1$ outgoing edges. This is $r^i$ possibilities at most. At the root, the number of outgoing edges is equal to $i$, so this is $r^i+1$. The total number of possibilities is then at most $r^{2i+1}$ since the sum of the
degrees is $2k$. Each point of the tree can have weight 0 or 1, which multiplies the expression by $2^{k+1}$.

**Proof of Theorem 1.1**: Let us consider each explanation tree a weighted tree in which the weight is 1 in a node exactly if the node is in Noise'. For each $n$, let $E_n$ be the set of possible explanation trees $ExpI$ for $u$ with weight $|F(ExpI)| = n$. First we prove the theorem for $m = \infty$, i.e. Noise' = Noise. If we fix an explanation tree $ExpI$ then all the events $w \in Noise'$ for all $w \in F = F(ExpI)$ are independent from each other. Therefore the probability of the event $F \subseteq Noise'$ is at most $\varepsilon^n$. Therefore

$$\Pr\{\xi(u) = 1\} \leq \sum_{n=1}^{\infty} |E_n| \varepsilon^n.$$  

By the Explanation Tree Lemma, each tree in $E_n$ has at most $k = 4n-4$ edges. Hence

$$\Pr\{\xi(u) = 1\} \leq 2r \cdot (2r^2)^{4n-4},$$

In the case $C \neq C'$ this estimate bounds only the probability of $\xi'(u) = 1$. $|ExpI(u, \xi')| \leq m$, since otherwise the events $w \in Noise'$ are not necessarily independent for $w \in F$. Let us estimate the probability that an explanation $ExpI(u, \xi')$ has $m$ or more nodes. It follows from the Tree Separator Theorem that $ExpI$ has a subtree $T$ with weight $n'$ where $m/12 \leq n' \leq m/4$, and at most $m$ nodes. Since $T$ is connected no two of its nodes can have the same projection. Therefore for a fixed tree of this kind, for each node of weight 1 the events that they belong to Noise' are independent. Hence for each tree $T$ of these sizes, the probability that $T$ is such a subtree of $ExpI$ is at most $\varepsilon^{m/12}$. To get the probability that there is such a subtree we multiply by the number of such subtrees. An upper bound on the number of places for the root is $tm^2 N$. An upper bound on the number of trees from a given root is obtained from the Tree Counting Lemma. Hence

$$\Pr\{|ExpI(u, \xi')| > m\} \leq 2tm^2 N \cdot (2r^{2} \varepsilon^{1/12})^m.$$

**3 The existence of small explanation trees**

**3.1 Some geometrical facts**

Three linear functionals are defined as follows for $v = (x, y, z, t)$.

$$L_1(v) = -x, \quad L_2(v) = -y, \quad L_3(v) = x + y.$$
Notice $L_1(v) + L_2(v) + L_3(v) = 0$. For a set $S$, we write

\[ \text{Size}(S) = \sum_{i=1}^{3} \max_{v \in S} L_i(v). \]

For every point $v$, $\text{Size}\{v\} = 0$.

A set $S = \{S_1, \ldots, S_n\}$ of sets is connected by intersection if the graph over $S$ is connected which we obtain by introducing an edge between $S_i$ and $S_j$ whenever $S_i \cap S_j \neq \emptyset$.

A spanned set is an object $P = (P, v_1, v_2, v_3)$ where $P$ is a space-time set, $v_i \in P$. The points $v_i$ are the poles of $P$, and $P$ is its base set. We define $\text{Span}(P)$ as $\sum_{i=1}^{3} L_i(v_i)$.

(3.1) **Spanned Set Creation Lemma**  If $P$ is a set then there is a spanned set $(P, v_1, v_2, v_3)$ on $P$ such that $\text{Span}(P) = \text{Size}(P)$.

**Proof:** We assign $v_i$ to a point of the set $P$ in which $L_i$ is maximal.

(3.2) **“Stokes’s” Theorem for spanned sets** Let $L = (L, u_1, u_2, u_3)$ be a spanned set and $\mathcal{M}$ be a set of subsets of $L$ connected by intersection, with $\{u_1, u_2, u_3\}$ contained in $\bigcup_{M \in \mathcal{M}} M$. Then there is a set $\{M_1, \ldots, M_k\}$ of spanned sets whose base sets are elements of $\mathcal{M}$, such that the following holds. Let $M'_i$ be the subset of $M_i$ consisting of its poles.

(a) $\text{Span}(L) = \sum_i \text{Span}(M'_i)$.

(b) Each $u_i$ is in one of the $M'_i$.

(c) The system $\{M'_1, \ldots, M'_k\}$ is a minimal system connected by intersection that connects the three poles $u_i$.

**Proof:** Without loss of generality, suppose that set $M_i \subset \mathcal{M}$ contains the point $u_i$. Let us choose $u_i$ as the $i$-th pole of $M_i$. Now leave only those sets of $\mathcal{M}$ that are needed for a spanning tree in the minimum tree connecting $M_1, M_2, M_3$. Keep deleting points from each set (except $u_i$ from $M_i$) until every remaining point is necessary for a connection among $u_i$. There will only be two- and three-element sets. Let us draw an edge between each pair of points if they belong to a common set $M'_i$. This turns the set $\bigcup_i M'_i$ into a graph. (Actually, this graph can have only two simple forms: a point connected via disjoint paths to $u_i$ or a triangle connected via disjoint paths to $u_i$.) For each $i$ and $j$, there is a shortest path between $M'_i$ and $u_j$. The point of $M'_i$ where this path leaves $M'_i$ will be made the $j$-th pole of $M_i$. (This rule puts three poles into each $M_i$.) Let us show that it puts the same number of each kind of pole into each point different from $u_1, u_2, u_3$ and therefore, the sum of spans did not change:

$$\sum_i \text{Span}(M'_i) = \text{Span}(L).$$  \hfill (3.3)
Let \( v \) be a point. Let \( H_1, \ldots, H_p \) be the sets \( M_v \) such that \( v \) is a pole of \( M_v \). For \( k = 1, \ldots, p \), let \( E_k \) be the set of those \( j \in \{1, 2, 3\} \) that the shortest path from \( v \) to \( u_j \) goes through \( H_k \). Then the sets \( E_k \) form a partition of the set \{1, 2, 3\}. Let \( E'_k = \{1, 2, 3\} \setminus E_k \). Then \( v \) is the \( j \)-th pole of \( H_k \) if and only if \( j \in E'_k \). Let \( \epsilon_k(j) = 1 \) if \( j \in E_k \) and 0 otherwise. Then \( j \in E'_k \) if and only if \( 1 - \epsilon_k(j) = 1 \). The number of \( k \)'s such that \( v \) is the \( j \)-th pole of \( H_k \) is therefore

\[
\sum_k (1 - \epsilon_k(j)) = p - \sum_k \epsilon_k(j) = p - 1.
\]

This number does not depend on \( j \), which is what we had to prove \( \blacksquare \)

### 3.2 Building an explanation tree

Let \( v = (a, b, u, t + 1) \) with \( \xi(v) = 1 \). If \( v \not\in Noise' \) then there is an \( \alpha \) such that \( \xi'(\alpha) = 1 \) for at least two members \( w \) of the set

\[
\{(a, b, \alpha, t), (a + 1, b, \alpha, t), (a, b + 1, \alpha, t)\}.
\]

Let us define the set \( Excuse(v) \) as such a pair of elements \( w \). We define \( Excuse(v) \) to be empty in all other cases. By Lemma 3.1, we can turn \( Excuse(v) \) into a spanned set, \( (Excuse(v), v_1, v_2, v_3) \) with span 1. Let us denote \( Excuse_i(v) = v_i \).

#### (3.4) Excuse Size Lemma

If \( V = (V, v_1, v_2, v_3) \) is a spanned set and \( \alpha \) are not in \( Noise' \) then

\[
\sum_i L_i(Excuse_i(v_i)) = \text{Span}(V) + 1.
\]

**Proof:** Let \( T \) be the triangle of points \( u \) satisfying the inequalities \( L_1(u) \leq 0 \)

\( L_2(u) \leq 0 \), \( L_3(u) \leq 1 \). Then \( T \) has size 1 and \( v + T \) always contains the set \( Excuse(v) \). Since the chosen poles turn \( Excuse(v) \) into a spanned set of size 1 the function \( L_1 \) achieves its maximum in \( T + v \) on \( Excuse_i(v) \). We have

\[
L_i(Excuse_i(v)) = \max_{u \in T + v} L_i(u) = \max_{u \in T} L_i(u) + L_i(v)
\]

Hence

\[
\sum_i L_i(Excuse_i(v_i)) = \sum_i \max_{u \in T} L_i(u) + \sum_i L_i(v_i)
\]

\( = \text{Size}(T) + \text{Span}(V) = 1 + \text{Span}(V). \)

\( \blacksquare \)

For every \( b \) with \( \xi'(b) = 1 \) we define \( V \) (the set of points that participated in “misleading” \( b \)) the sets \( Arrows \) and \( Forks \) of undirected edges on \( V \) as follows. \( V = V_0 \), \( Arrows = A_n \), \( Forks = F_n \), where the sets \( V_i, A_i, F_i \) (\( i = 0, \ldots, n \)) and the number \( n \) are defined as follows. Let \( V_0 = \{b\} \), \( A_0 = F_0 = \emptyset \). Suppose that \( V_i, A_i, F_i \) is defined. If for all \( v \in V_i \) with \( \xi(v) = 1 \) we have \( Excuse(v) \subset V_i \) then \( n = i \). Else let \( v \) be such that \( \xi(v) = 1 \) and \( Excuse(v) \not\subset V_i \). Let
Excuse(\(v\)) = \{a, b\}. Then
\[
\begin{align*}
V_{i+1} &= V_i \cup \{a, b\}, \\
A_{i+1} &= A_i \cup \{\{v, a\}, \{v, b\}\}, \\
F_{i+1} &= F_i \cup \{\{a, b\}\}.
\end{align*}
\]

For an arbitrary node \(u\) of the investigation graph defined above, \(G_u\) is a subgraph of \((V, \text{Arrows})\) induced by
\[
\{v \in V : \text{Time}(v) \leq \text{Time}(u)\}.
\]

Notice that \(G_u\) includes points \(v\) with \(\text{Time}(v) = \text{Time}(u)\). The set \(\text{History}(u)\) is the connected component of \(G_u\) containing \(u\). We will use the notation \(\text{Time}(K)\) for histories \(K\) without ambiguity. For a history \(K\) we define the graph \(G_K = (V_K, E_K)\) as follows:
\[
\begin{align*}
V_K &= \{\text{history } R \subseteq K : \text{Time}(R) = \text{Time}(K) - 1\}, \\
\{R, S\} \in E_K \text{ iff } \{R, S\} \subseteq V_K \text{ and for some } \{v, w\} \in \text{Forks} \text{ we have } v \in R \text{ and } w \in S \text{ and } \text{Time}(v) = \text{Time}(w) = \text{Time}(K) - 1.
\end{align*}
\]

(3.5) Lemma \(G_K\) is connected.

Proof: In the graph of arrows the history \(K\) is a connected component. The subhistories in \(G_K\) are connected with each other only through pairs of arrows going through time \(\text{Time}(K)\). The tails of each such pair of arrows are connected by a fork \(\square\).

A spanned history is a spanned set that is a history in which all spanning points \(v_i\) have the maximum value of \(\text{Time}\). It is easy to see that a spanned history goes through several instants of time if and only if none of its spanning points is in \(\text{Noise}'\).

The explanation tree will be built from an intermediate object that is a tree whose nodes are of two kinds: spanned histories, and intermediate nodes. The edges of the tree are original edges of the graph adjacent to the intermediate nodes or the poles of these spanned histories. Such a tree will be called a partial explanation tree. The span of the explanation tree will be the sum of the spans of its nodes (intermediate nodes have span 0) and the sizes of its forks (horizontal edges: their sizes are all 1).

We start with a node \(u \notin \text{Noise}'\) such that \(\xi'[u] = 1\). Then \((\text{History}(u), u, u, u)\) is itself a spanned history, forming a one-node partial explanation tree. Now we apply repeatedly an operation called refinement to the tree.

Let \(K = (K, \text{v}_1, \text{v}_2, \text{v}_3)\) be a spanned history in the tree going through several instants of time. Then \(\text{v}_i\) are not in \(\text{Noise}'\). Consider the graph \(G_K = (V_K, E_K)\) defined above. Let \(\mathcal{M} = V_K \cup E_K\), i.e. the set of all histories in \(V_K\) and all
edges in $G_K$ connecting them, taken as two-element sets. Let $L$ be the union of these sets, $L = (L, u_1, u_2, u_3)$ where $u_i = \text{Excise}(v_i)$. Lemma 3.5 implies that the set $M$ is connected by intersection. Let us apply Lemma 3.2 to $L$ and $M$. We find a set $M_1, \ldots, M_n$ of spanned sets such that

$$\sum_i \text{Span}(M_i) = \text{Span}(L) = \sum_i L_i(u_i).$$

It follows from Lemma 3.4 that the latter sum is $\text{Span}(K) + 1$. Also, the $u_i$ are among the poles of these sets. Some of these sets are disjoint spanned histories, others are forks connecting them, adjacent to their poles. Consider these forks again as edges and the spanned histories as nodes. By the minimality property of the lemma, they form a tree. Now the refinement operation deletes all non-pole nodes of the spanned set $K$ in the old tree and adds the tree just built. For each of the 1, 2 or 3 nodes that were poles in $K$ it chooses a pole $v_i$ in it and adds the arrow from $v_i$ to $u_i$ to connect the new tree to the old one. Notice that the operation increased the span by at least 1 and the number of arrows by at most 3.

When the refinement operation cannot be applied any longer then all nodes of the tree belong to $\text{Noise}^e$.

Proof of Lemma 2.1: What is left to prove is the estimate on the number of edges. Let us contract each arrow $(u, v)$ of the explanation tree one-by-one into its bottom point $v$. The edges of the resulting tree are the forks. All the intermediate nodes will be contracted into the remaining unrefinable one-node spanned histories that are actually elements of $\text{Noise}^e$. If $k$ is the number of these nodes then $k - 1$ is the number of forks. The span of the explanation tree just constructed is the sum of sizes of the forks, i.e. $k - 1$. The number of arrows is at most $3(k - 1)$ since each introduction of 3 arrows was accompanied by an increase of the span by 1. The total number of edges of the explanation tree is thus at most $4(k - 1)$. \qed

References

[1] Piotr Berman and Janos Simon. Investigations of fault-tolerant networks of computers. In Proc. of the 20-th Annual ACM Symp. on the Theory of Computing, pages 66 – 77, 1988.

[2] Peter Gács and John Reif. A simple three-dimensional real-time reliable cellular array. Journal of Computer and System Sciences, 36(2):125–147, April 1988.

[3] Andrei L. Toom. Stable and attractive trajectories in multicomponent systems. In R. L. Dobrushin, editor, Multicomponent Systems, volume 6 of Advances in Probability, pages 549–575. Dekker, New York, 1980. Translation from Russian.