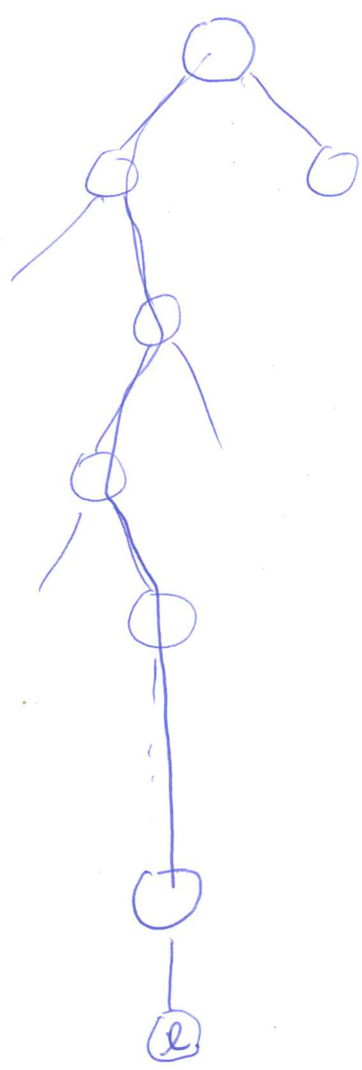
# Pumping lemma for context-free languages

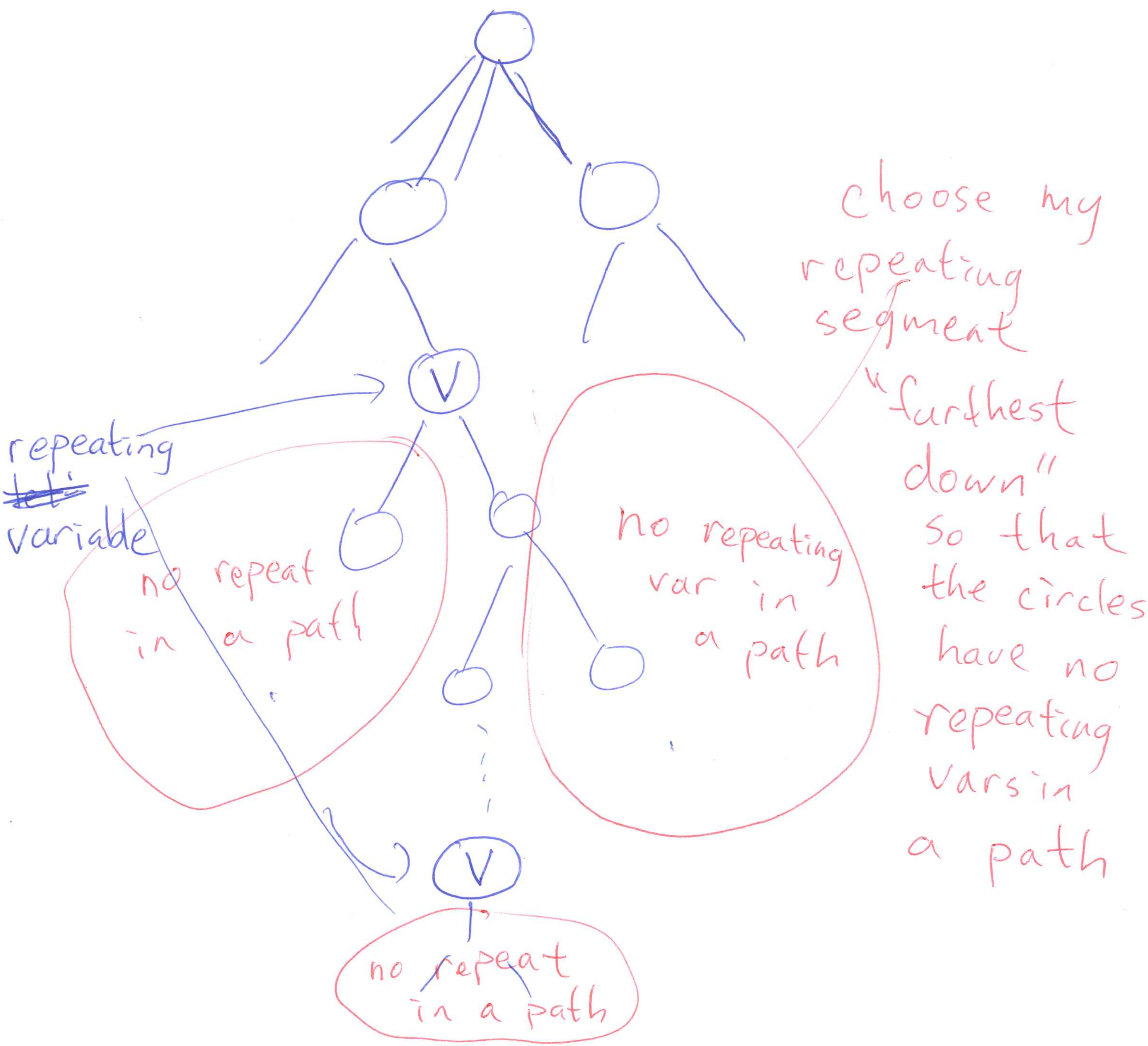
10/29

~~We~~ Suppose we have a context-free language $L$. It has a grammar $G$ which I'll assume is in Chomsky Normal Form. $G$ has some number of variables $\alpha = |V|$. Look at a derivation tree for $w$ really long ($2^\alpha$ is long enough) $\Rightarrow$ this means my tree is tall.

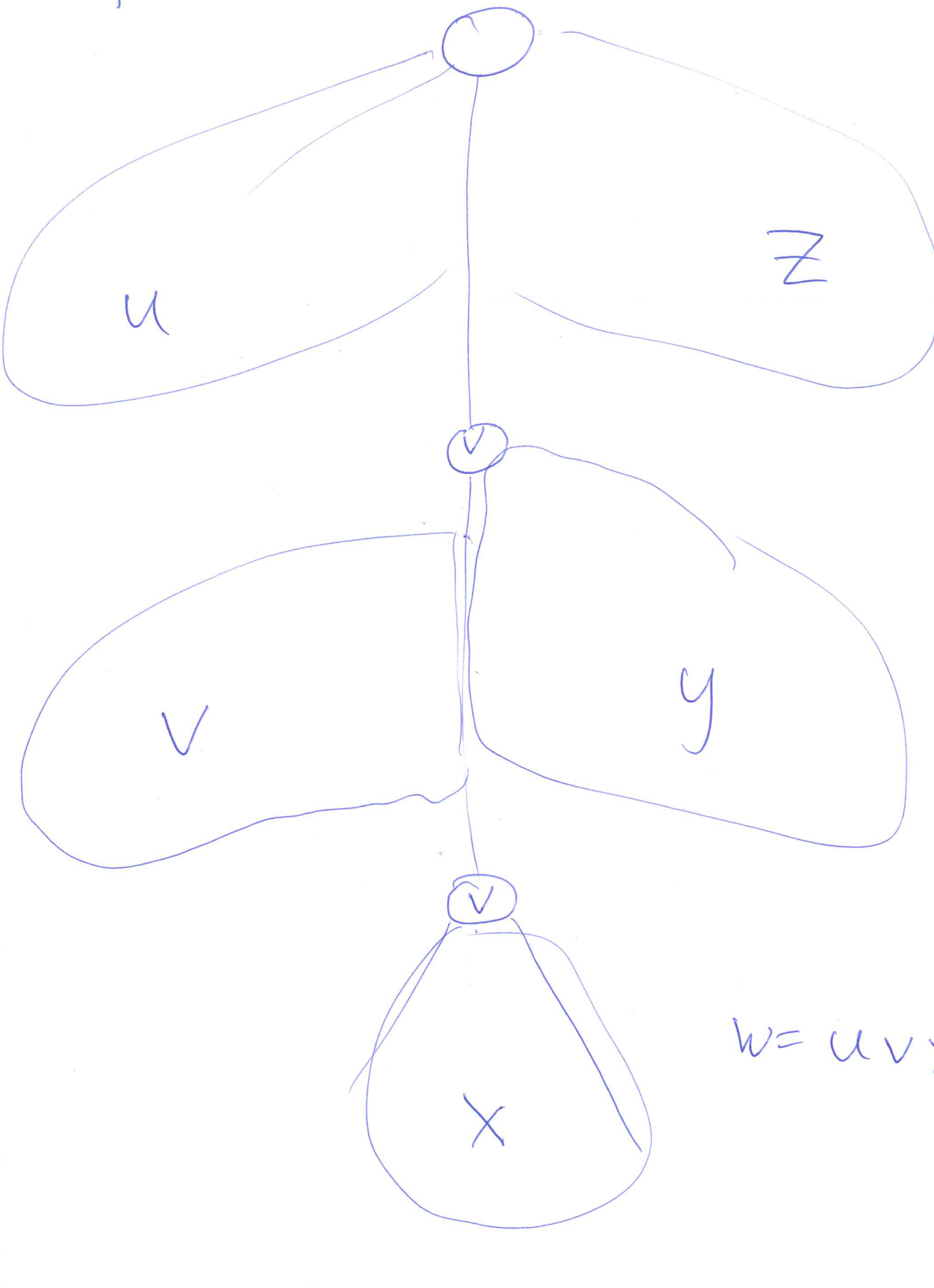


This tree is tall, so some path through the tree has $\geq \alpha$ internal nodes.

The pigeonhole principle tells me 2 of the nodes on this path are labelled by the same variable. So we can repeat the segment between the repeating variable as much as we want.



Call the parts of the string given by the tree




$w = uvxyz$

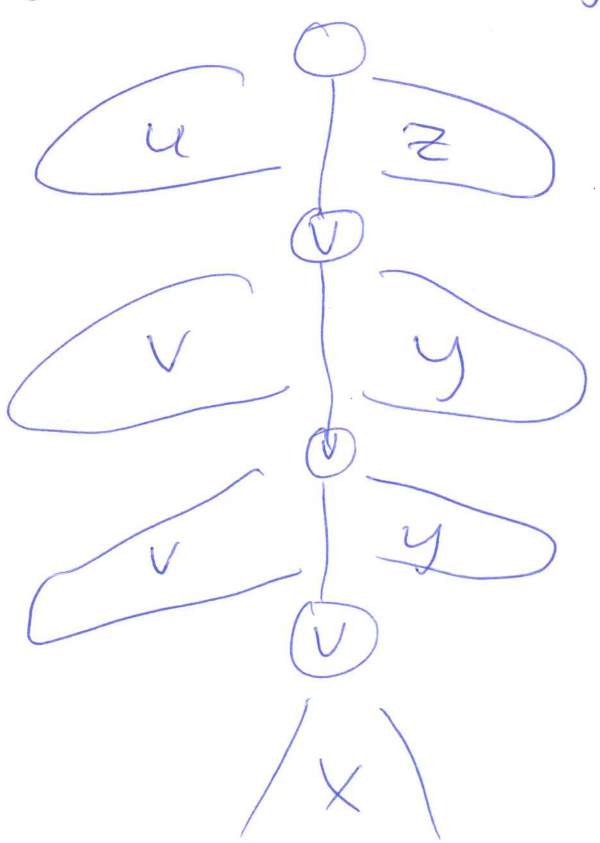
I can repeat both v and y (as long as I repeat both the same number of times)

$|v| \le a2^a$ $|y| \le a2^a$ $|x| \le 2^a$

Let $m = a2^a + a2^a + 2^a$

So $|vxy| \le m$. (It really doesn't matter what m is - just that it only depends on G, not on w)

$uv^ixy^iz$ is in my language b/c




is a derivation tree.

**Pumping Lemma:** Given a CFL $L$, there exists a number $m$ such that, for every $w \in L$, there exists $u, v, x, y, z$, $w = uvxyz$, $|vxy| \le m$, $v \ne \lambda$ or $y \ne \lambda$ such that

$$uv^i x y^i z \in L \text{ for all } i.$$

**Prop:** $L = \{a^n b^n c^n \mid n \in \mathbb{Z}\}$ is not a context-free language

**Pf:** Given $m$, pick $w = a^m b^m c^m$.

Since $|vxy| \le m$, a....ab|....bc....c

$vxy$ can't have both $a$'s and $c$'s

We have several cases for the contents of $v$ & $y$:

Case 1: $vy = a^k$

Case 2: $vy = a^k b^\ell$

Case 3: $vy = b^k$

Case 4: $vy = b^k c^\ell$

Case 5: $vy = c^k$

(In no case can we have all 3 ~~let~~ letters in $vy$ since $|vxy| \le m$)

In all of these cases

$uv^2xy^2z \notin L$, since it won't have equal numbers of a's, b's & c's.

This contradicts the pumping lemma, so $L$ can't be context-free.

---

Reminder on using pumping lemma for contradiction proof:

We don't get to choose $m$ – we need to handle every possibility

We do get to choose $w$ (depending on $m$)

We don't get to choose $u, v, x, y, z$ – we need to handle every possibility with $|vxy| \leq m$, $v \neq \lambda$ or $y \neq \lambda$.

We do get to choose $i$ = how many times $v$ & $y$ are repeated.

$L = \{ww \mid w \in \{a,b\}^*\}$ is not ~~context-fre~~ context-free.

---

Pf: Given m, pick

$$w = a^m b^m a^m b^m$$

$a \cdots\cdots ab \cdots\cdots b\, a \cdots\cdots ab \cdots\cdots b$

If $\{vxy\}$ is all in the left half, then, in $w_2 = uv^2xy^2z$, the midpoint of the string will have moved left, so $w_2$ is not in L as the 2nd half begins w/ b's and the first half begins with a's. (or, the midpoint moved all the way into the a's, in which case the left half has no b's but the right half does)

If vxy is all in the right half, same.

If $vxy$ is in the middle, then either $v$ is all b's or $y$ is all a's (and, also $v$ starts w/ a b & $y$ ends w/ an a)
Then $w_2 = uv^2xy^2z$ has more consec b's in the left half than the right half (if $v$ is all b's) or more consec a's in the right half than the left half (if $v$ is all a's).