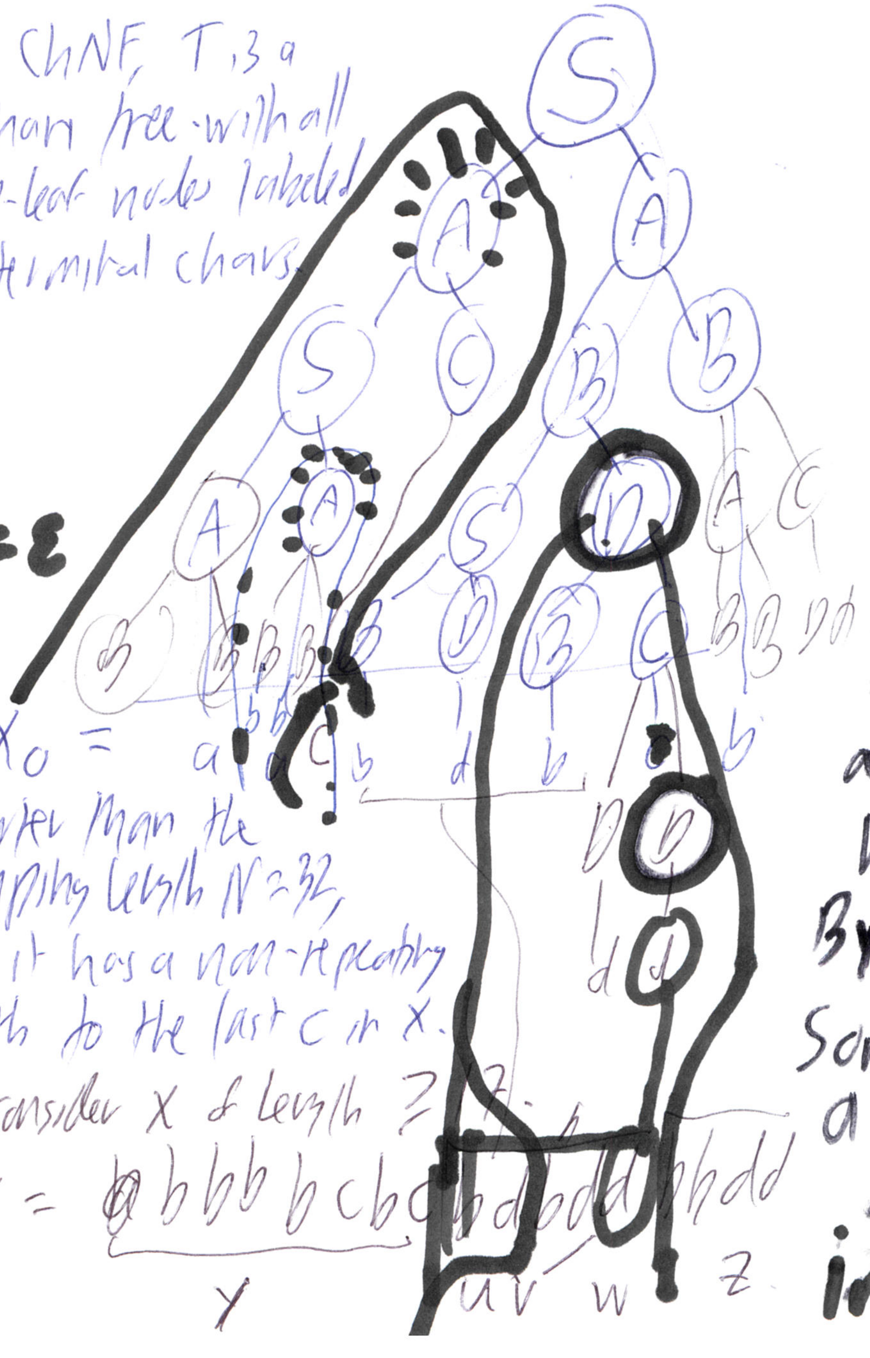
Suppose we have a CFG $G = (V, \Sigma, R, S)$ in Chomsky NF. Set $k = |V|$, $N = 2^k$. Consider any $x \in L(G)$ with $|x| > N$ and take a parse tree T for x from S.

By ChNF, T is a binary tree with all non-leaf nodes labeled by terminal chars.

$S \to AA \mid BD \mid$
$A \to SC \mid BB \mid a$
$B \to AC \mid SD \mid b$
$C \to AB \mid DD \mid c$
$\qquad\qquad\qquad \mid d$
$D \to BC$



$y = \varepsilon$

$X_0 =$ a a c b d b

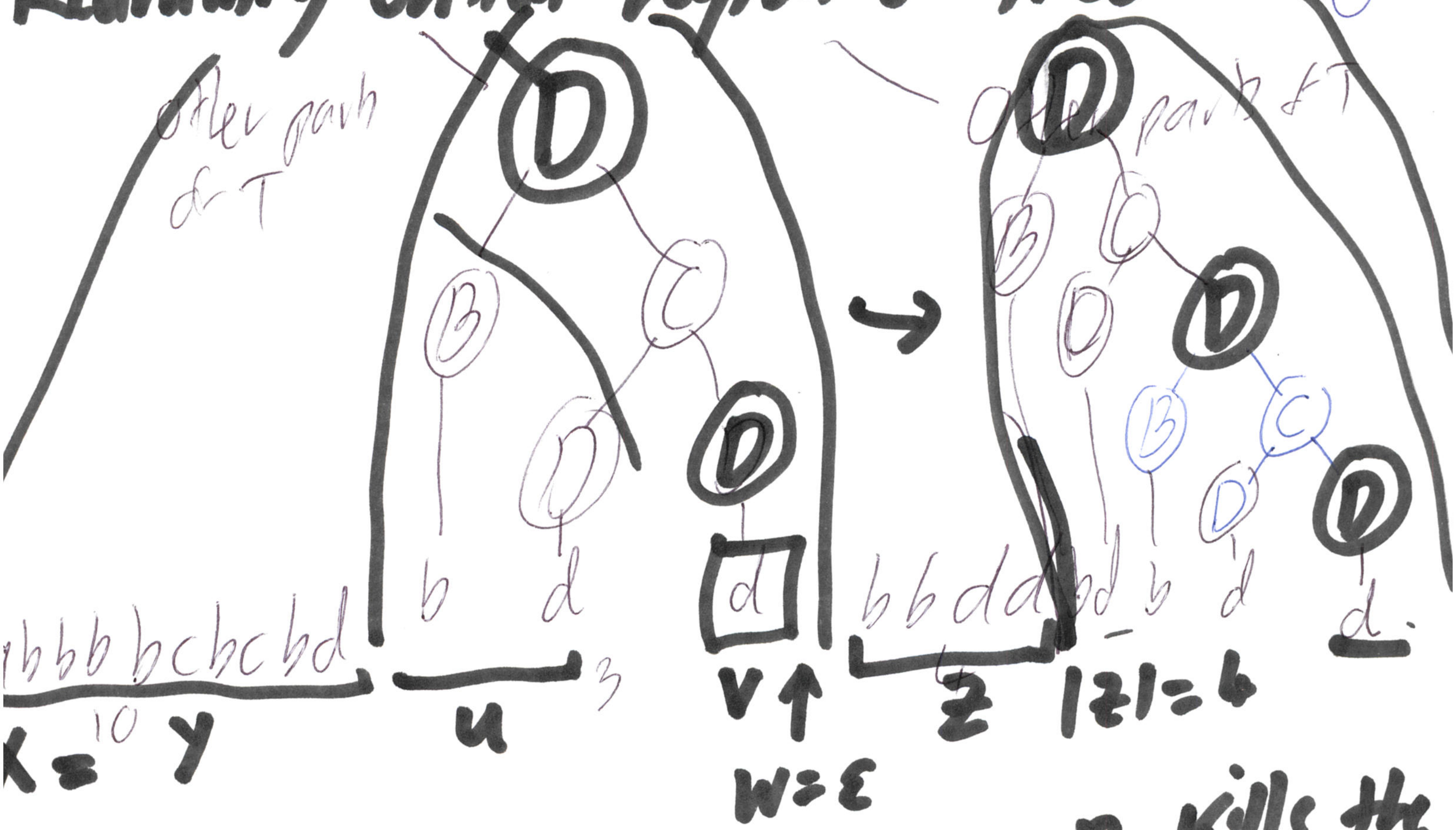
shorter than the pumping length $N = 32$, and it has a non-repeating path to the last c in X.

But consider X of length ≥

$X = \underbrace{abbbbcbc}_{x}\, \underset{u}{b}\,\underset{v}{d}\,\underset{w}{b}\,\underset{z}{dd}\,bdd$

Key Fact: By $N = 2^k$ if $|X| > N$ then T has a path with at least $k+1$ internal nodes. By Pigeonhole Principle some variable repeats along that path. Take a repeated variable D in the bottom $\underline{k+1}$ nodes

Redrawing critical region of tree



Idea: • Move lower D to upper D. Kills the B, C, D nodes.

Get: $x' = yvz$ = abbbbcbcbd·$\underline{d}$·bbdd

• Repeat the lower D as the upper D for $i = 2$ times. Get $x^{(2)}$ = abbbbcbcbd·$\underbrace{\text{bdbd}}_{\text{repeated } i=2 \text{ times}}$·$\underline{d}$·bbdd.

Summary:

I show with $w = \varepsilon$

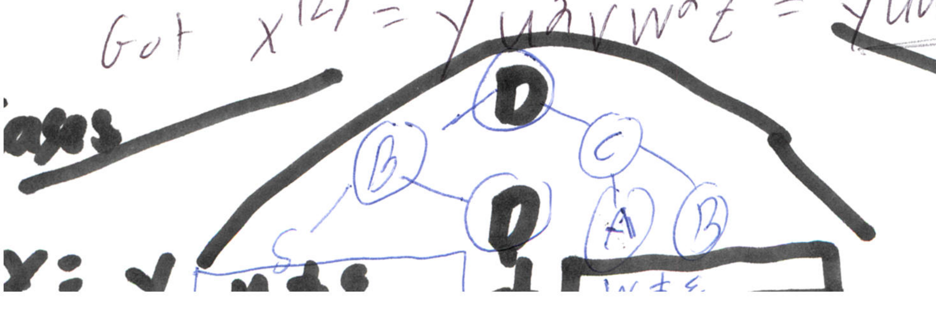
$\boxed{\text{Always at least one of } u \text{ and } w \text{ must be nonempty}}$

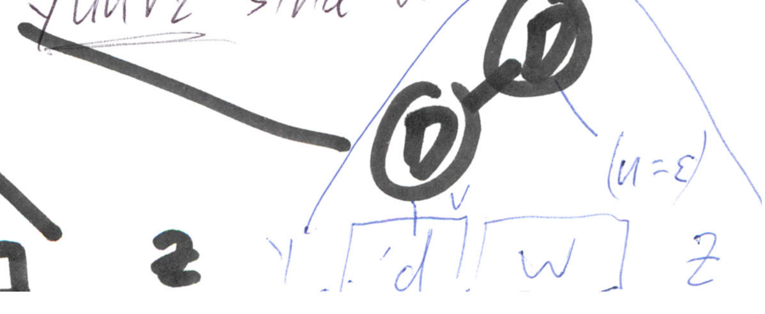
Original $x = yuvwz$ Call this $x^{(0)}$

Got $x' = yu^0vw^0z = \underline{yvz}$

Got $x^{(2)} = yu^2vw^2z = \underline{yuuvz}$ since $w = \varepsilon$

Cases





Theorem: Given any context-free language L, there are ~~numbers~~ is a number $K, N > 0$ such that for all $x \in L(G)$, $|x| > N$, we can break $\equiv \exists y,u,v,w,z$

$x =: yuvwz$ s.t. $(\forall i \geq 0)$, the string

$$x^{(i)} = yu^i v w^i z$$ belongs to L, and

where $|uvw| \leq N$, and $uw \neq \varepsilon$. $G = (V, \Sigma, R, S)$

Proof: By L being a CFL, L has a grammar G in ChNF. Take $K = |V|$ and $N = 2^K$. The breakdown follows as above. ■

Contrapositive: Suppose for all $N > 0$ there exists $x \in L(G)$ with $|x| > N$ s.t. for all breakdowns $x = yuvwz$ where $|uvw| \leq N$ and u, w are not both ε, there exists $i \geq 0$ such that $x^{(i)} = yu^i v w^i z \notin L$.

$\boxed{\text{Then } L \text{ is not a CFL.}}$

This sheet yields a "proof script" for proving that certain languages L are not context-free.

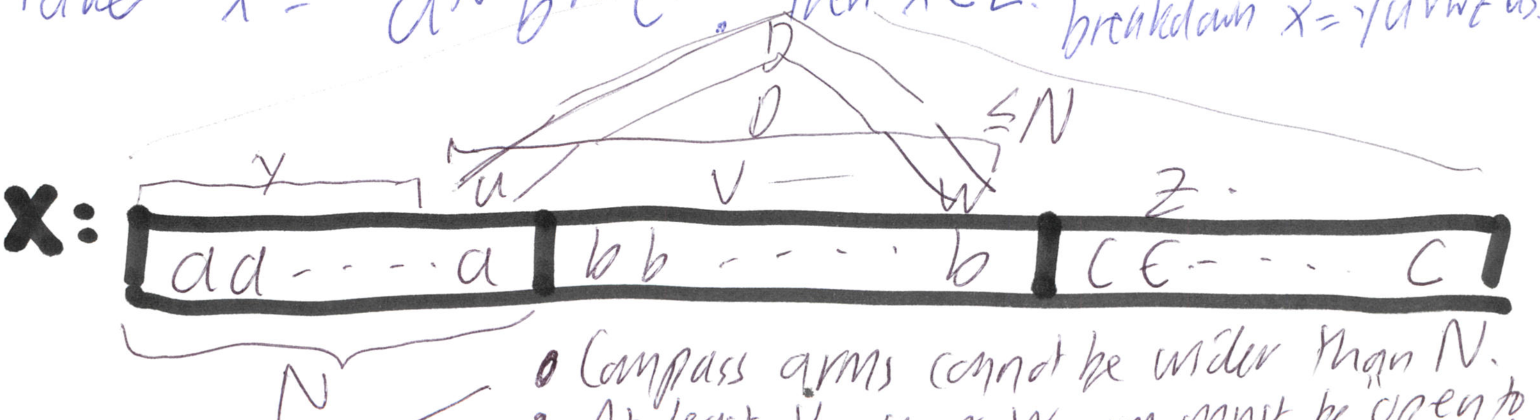
Let any $N > 0$ be given. Take $x =$ ____.

Consider any possible breakdown $x = yuvwz$ subject to $|uvw| \le N$ and u, w not both ε.

Take $i =$ ____. Then $x^{(i)} = yu^ivw^iz$ = ____ does not belong to L because ☆ ____.

Since N and the breakdowns are arbitrary, L is not a CFL, by the CFL Pumping Lemma.

Example: $L = \{a^n b^n c^n : n \ge 1\}$. Let $N > 0$ be given. Take $x = a^N b^N c^N$. Then $x \in L$. Visualize any possible breakdown $x = yuvwz$ as:



- Compass arms cannot be wider than N.
- At least the u or w arm must be open to generate.

Hence the compass cannot keep all of the a's, b's, c's in balance.
Hence $x^{(0)} = yvz$ subtracts at least one a, b, or c.
$\therefore$ $x^{(0)} \notin L$ because we can't keep $\#a = \#b = \#c$ in it. So L is not a CFL. ∎