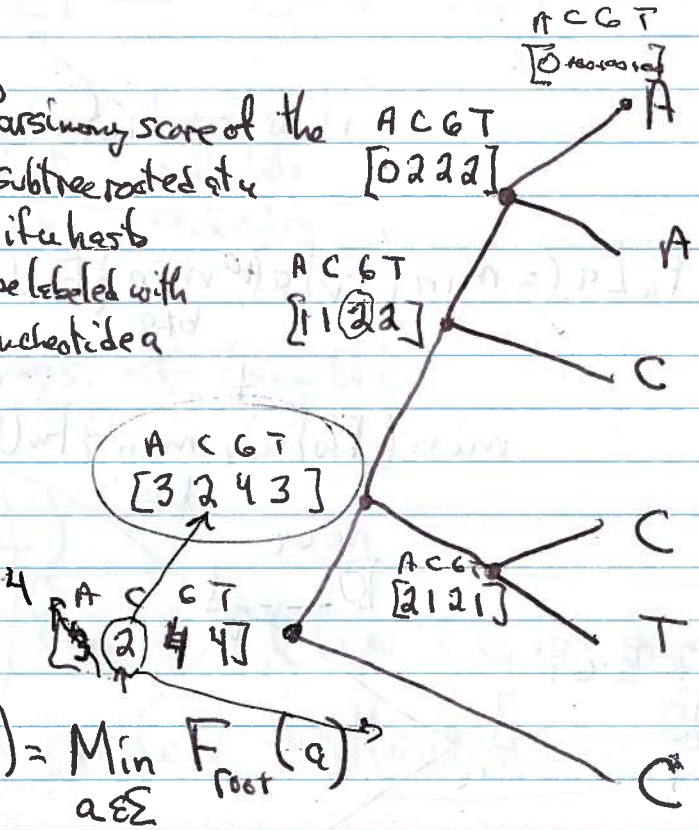
Sankoff Algorithm (1975)

Input: $\{S_1, S_2 \ldots S_n$ each of length 1
Tree T

Goal: Calculate ParsScore$(S_1 \ldots S_n, T)$

Define $F_u[a]$ = Parsimony score of the subtree rooted at u if u has b be labeled with nucleotide a

(u = node, $a \in \{A, C, G, T\}$)



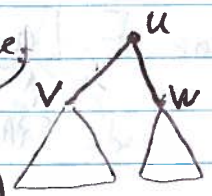
$$\text{ParsScore}(\text{Tree}) = \min_{a \in \Sigma} F_{\text{root}}(a)$$

Algo: For each node u in tree (from leaves back to the root)

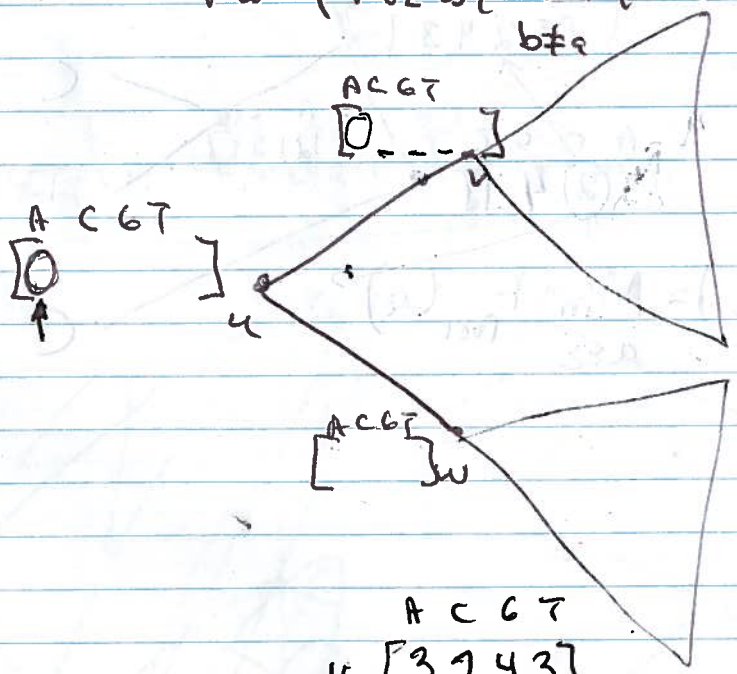
For each $a \in \{A, C, G, T\}$

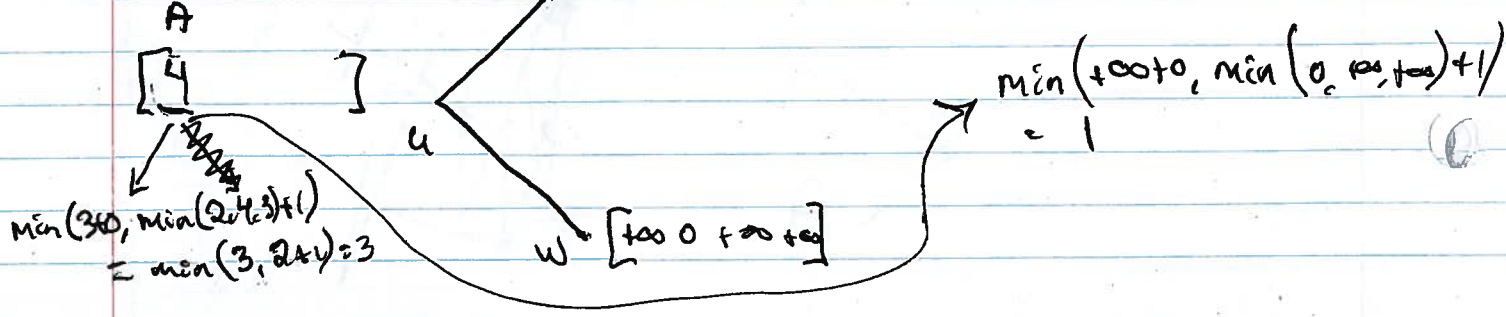
if u is a leaf $F_u[a] = \begin{cases} 0 \text{ if } s_u = a \\ +\infty \text{ if } s_u \neq a \end{cases}$

if v is an internal node,



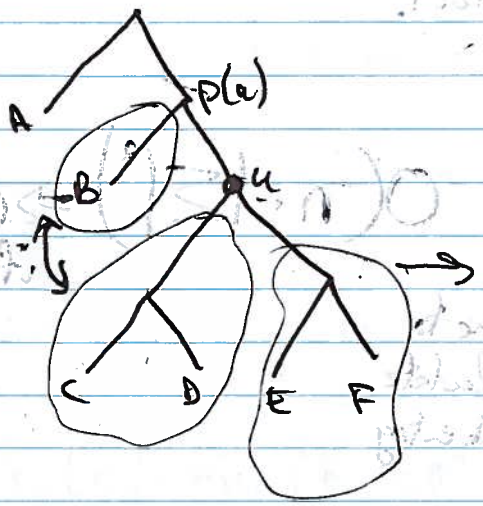
$$F_u[a] = \min\left(F_v[a], \min_{b \neq a}\{F_v(b)\} + 1\right) + \min\left(F_w[a], \min_{b \neq a}\{F_w(b)\} + 1\right)$$





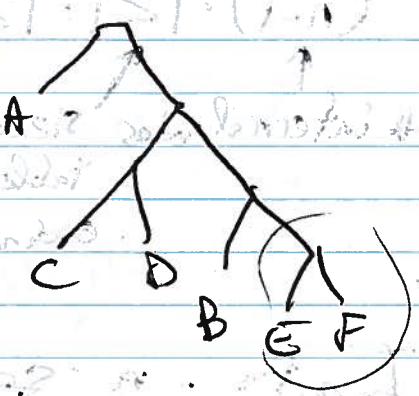
Tree neighborhood definitions

T:



Nearest-neighbor interchange (NNI)

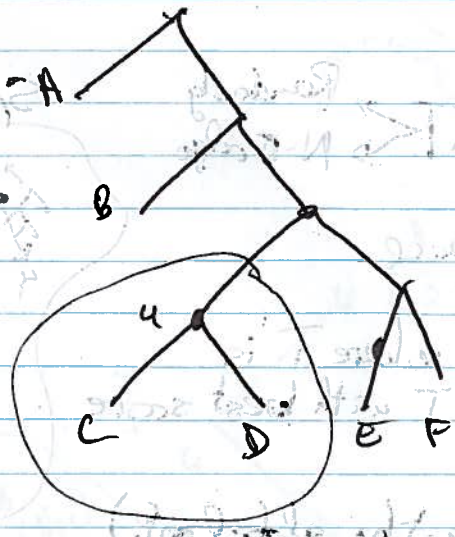
Pick some internal node u



$$Neighbors(T) = \{ T' \text{ s.t. } T' \text{ can be obtained from } T \text{ with } \underline{\text{one}} \text{ NNI} \}$$

$$|Neighbors(T)| = (n-2) \cdot 2$$

Subtree Prune and Regraft



Choose any node in T
- cut in off from T
- reinsert along other edge of T

⇒