Goals

- Prove correctness of Huffman's algorithm

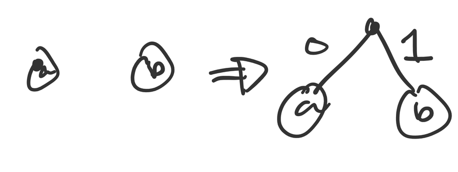
CCI Interview Prep Program: Handshake

Correctness: Huffman's Alg Always Produces a Prefix-free code with smallest possible average length

Proof:

- Prefix-free $\Longrightarrow$ Because of merging approach ✓ (Regular)
- Minimum Av. Length $\Longrightarrow$ induction on $|\Sigma| = n$.

Base case: If $n = 2$, two letters: a, b



$\Rightarrow$ Av Length 1 $\Rightarrow$ optimal

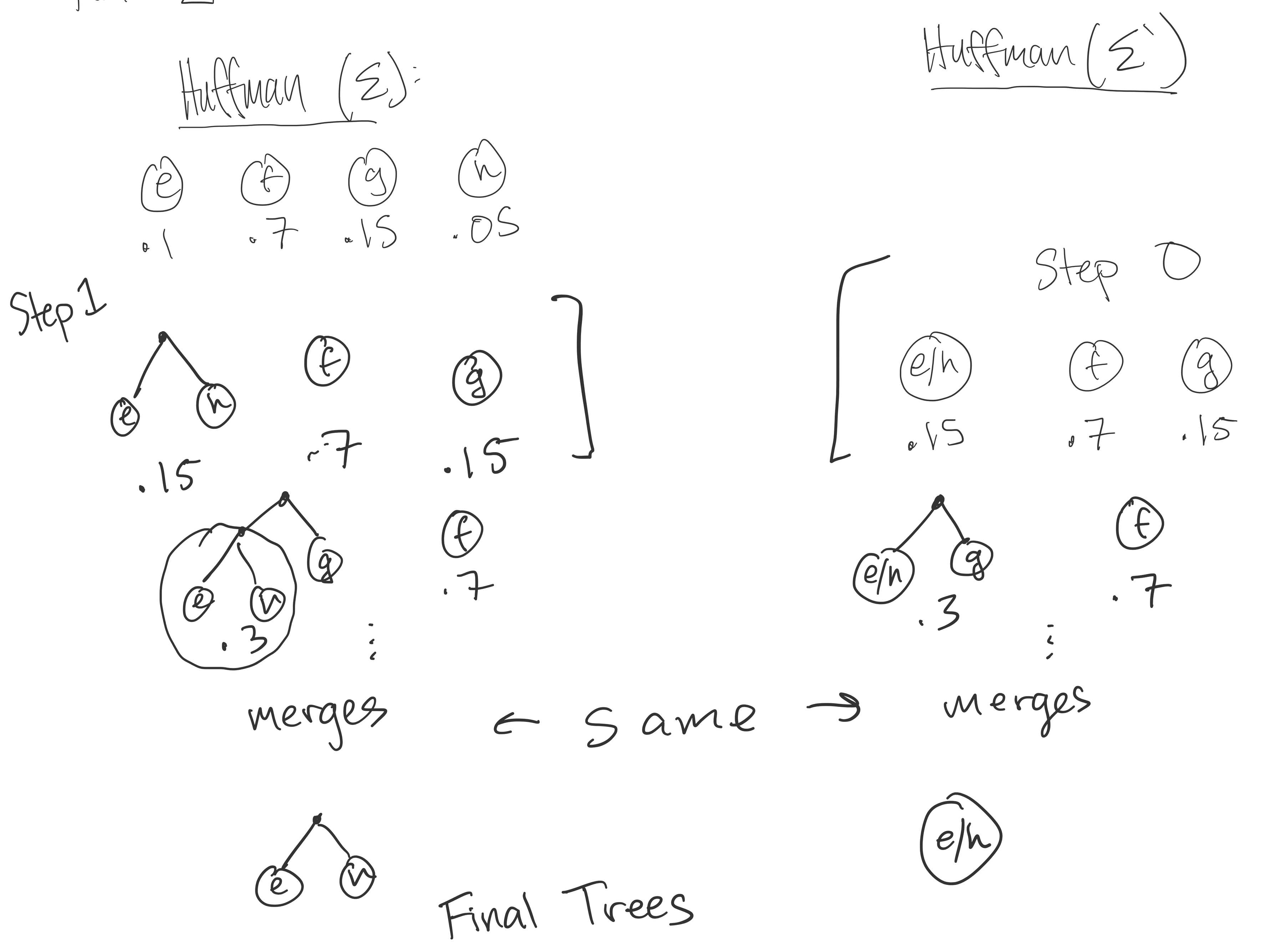
Inductive step: Let $k \geq 2$. Assume Huffman produces a code with smallest possible av. length for any input with $|\Sigma| = k$.

Consider an input with $|\Sigma| = k+1$. Let a, b be letters with the smallest probability. Let $\Sigma^- = \Sigma - \{a, b\} \cup \{a/b\}$, and $p(a/b) = p(a) + p(b)$.

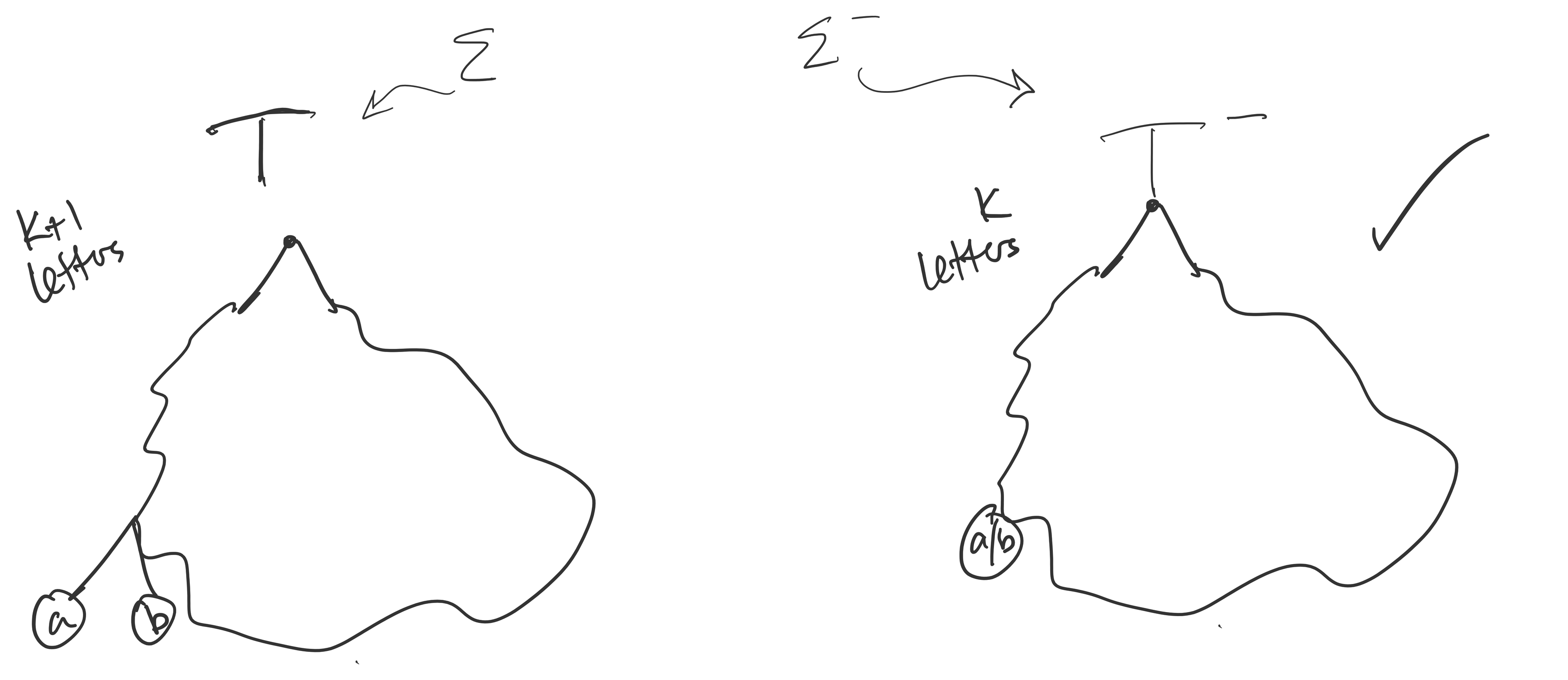
ex:

If $\Sigma = \{e, f, g, h\}$ $p(e) = .1$ $p(f) = .7$ $p(g) = .15$ $p(h) = .05$

then $\Sigma^- = \{e/h, f, g\}$ $p(e/h) = .15$ $p(f) = .7$ $p(g) = .15$



Final Trees

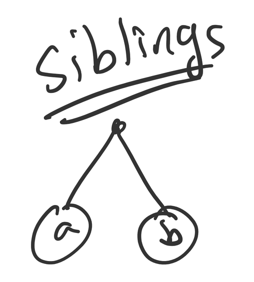


Since $|\Sigma^-| = k$, by inductive assumption T^- gives an optimal code (smallest av length).

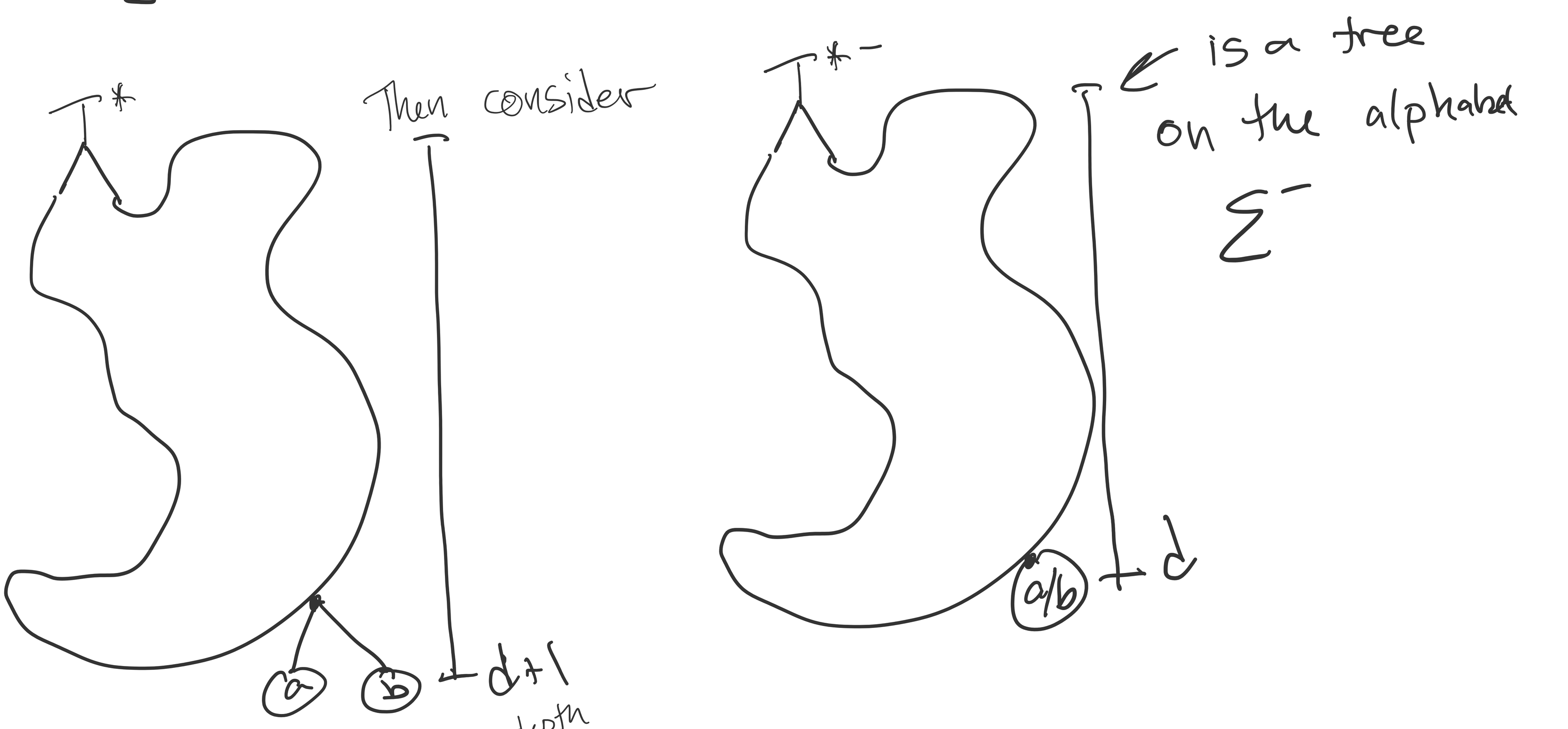
For contradiction, assume T is not optimal

> Lemma 1: There is an optimal tree with a, b siblings (Obviously Huffman's alg makes a, b siblings, but maybe there is a better approach that doesn't) [will prove later].

siblings



Let T^* be an optimal tree $(T \neq T^*)$, with a, b as siblings



Then:

$$L(T^*) = \sum_{i \neq a,b} p(i)\, d(i) + \boxed{p(a)(d+1) + p(b)(d+1)}$$ ✓

depth

$$L(T^{*-}) = \sum_{i \neq a/b} p(i)\, d(i) + \boxed{(p(a)+p(b))\, d}$$ ✓

$p(a/b)$

(explain notation)

$< \; = \; > \; ?$

so $L(T^*) - L(T^{*-}) \boxed{=} p(a) + p(b)$

Similarly, $L(T) - L(T^-) \boxed{=} \boxed{p(a) + p(b)}$

So $L(T^*) - L(T^{*-}) \boxed{=} L(T) - L(T^-)$

Rearranging: $\underbrace{L(T^-) - L(T^{*-})}_{} \boxed{=} \underbrace{L(T) - L(T^*)}_{\text{positive b/c } L(T) \text{ worse (larger) than } L(T^*)}$

This is a contradiction, because not positive (neg or 0) b/c T^- is optimal by our inductive assumption

so then $\boxed{T \text{ is optimal.}}$

Lemma 1: There is an optimal tree for Σ with a, b siblings.