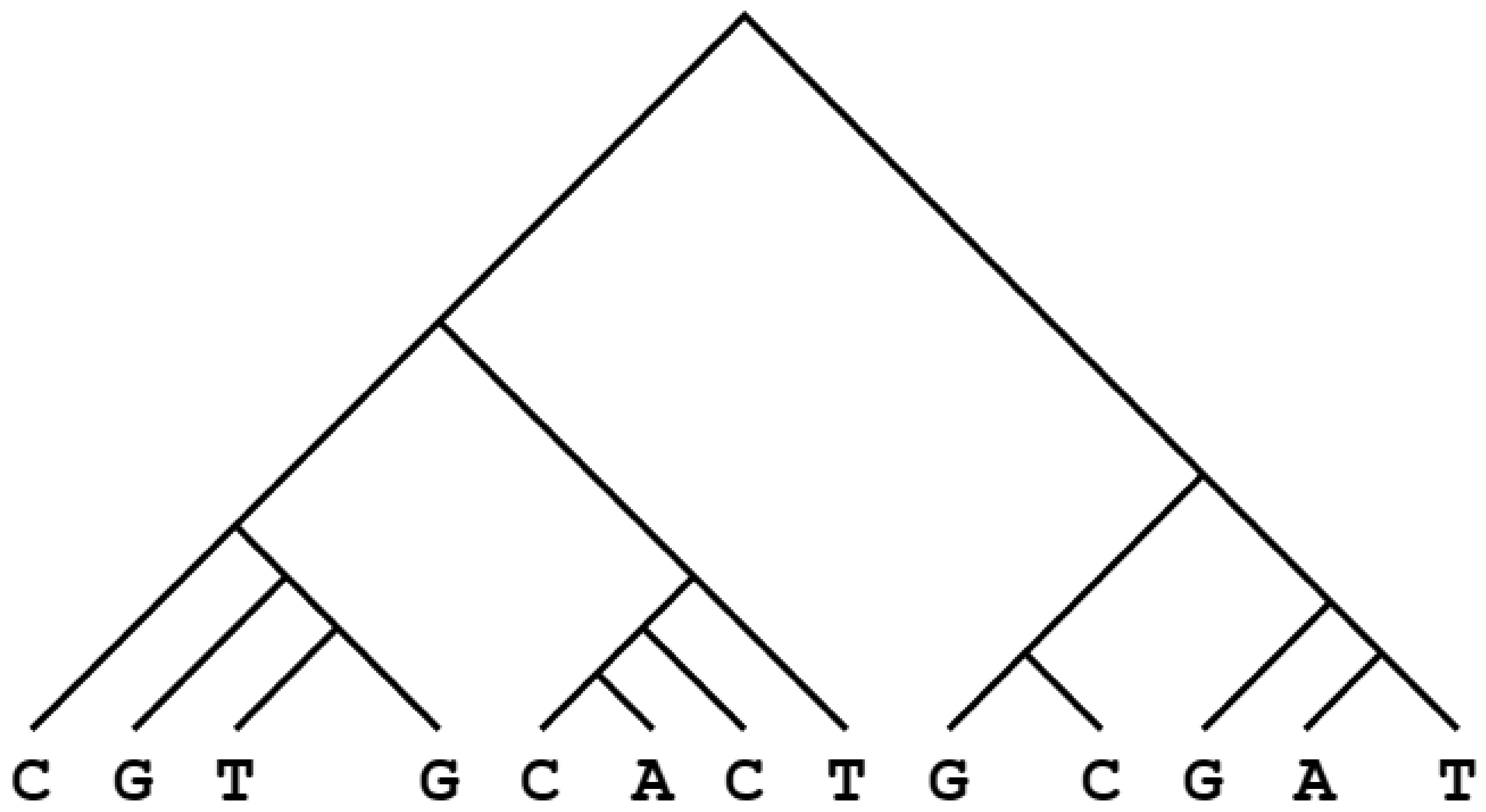
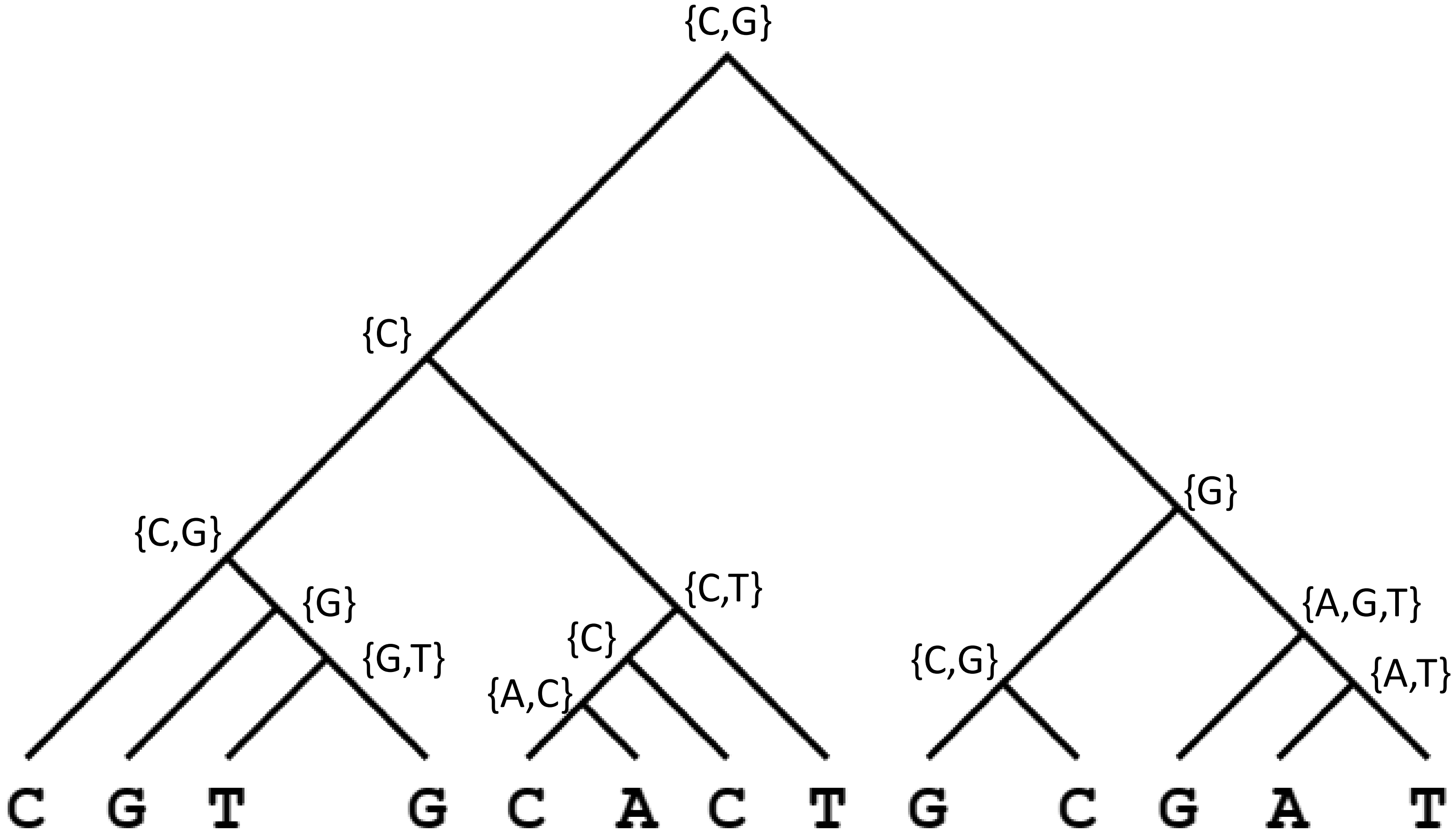
19. Übung

Einführung in die Bioinformatik I, 2. Teil
Sommersemester 2021

Aufgabe 1 (5 Punkte): Verwenden Sie den Fitch-Algorithmus, um eine maximal sparsame Benennung der inneren Knoten des unten gegebenen Baumes zu finden. Gibt es mehrere optimale Lösungen? Wenn ja, wie viele?

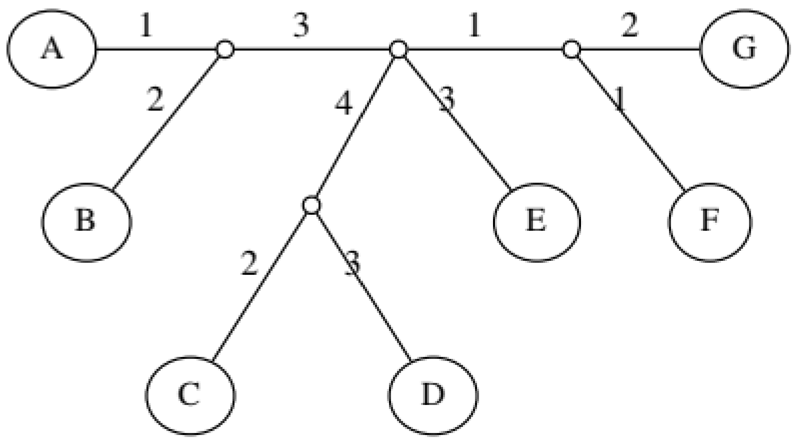


Aufgabe 1 (5 Punkte): Verwenden Sie den Fitch-Algorithmus, um eine maximal sparsame Benennung der inneren Knoten des unten gegebenen Baumes zu finden. Gibt es mehrere optimale Lösungen? Wenn ja, wie viele?



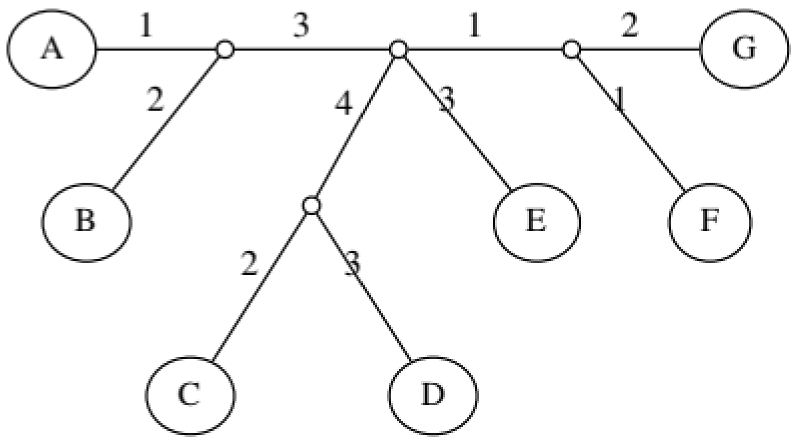
→ Es gibt insgesamt 4 optimale Lösungen

Aufgabe 2 (5 Punkte): Geben Sie die zugehörige Distanzmatrix für den dargestellten additiven Baum an.



A	B	C	D	E	F	G	
							A
							B
							C
							D
							E
							F
							G

Aufgabe 2 (5 Punkte): Geben Sie die zugehörige Distanzmatrix für den dargestellten additiven Baum an.



A	B	C	D	E	F	G	
0	3	10	11	7	6	7	**A**
	0	11	12	8	7	8	**B**
		0	5	9	8	9	**C**
			0	10	9	10	**D**
				0	5	6	**E**
					0	3	**F**
						0	**G**

Aufgabe 3 (5 Punkte): Beweisen Sie, dass für jede *beliebige* Metrik $d : X \times X \to \mathbb{R}$ gilt: $d(x, y) \geq 0$ für alle $x, y \in X$.

Angenommen es gibt zwei Elemente x_1, x_2 in einer Metrik mit $d(x_1, x_2) < 0$.

Da für alle Elemente $x, y, z \in X$ einer beliebigen Metrik gilt, dass

$$d(x, y) \leq d(x, z) + d(z, y),$$

dann muss auch folgendes gelten:

$$\underbrace{d(x_1, x_1)}_{\substack{\text{Metrik Regel 1:} \\ d(x,x) = 0}} \leq \underbrace{d(x_1, x_2)}_{\substack{\text{Annahme oben:} \\ d(x_1, x_2) < 0}} + \underbrace{d(x_2, x_1)}_{\substack{\text{Metrik Regel 2:} \\ d(x_2, x_1) = d(x_1, x_2) < 0}}$$

Da die Summer zweier negativer Zahlen nicht kleiner oder gleich 0 sein kann, gilt die Ungleichung nicht.
→ Widerspruch zur Annahme
→ $d(x, y) \geq 0$ für alle $x, y \in X$ einer beliebigen Metrik.
→ q.e.d

<u>Revision Übung 17:</u>

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

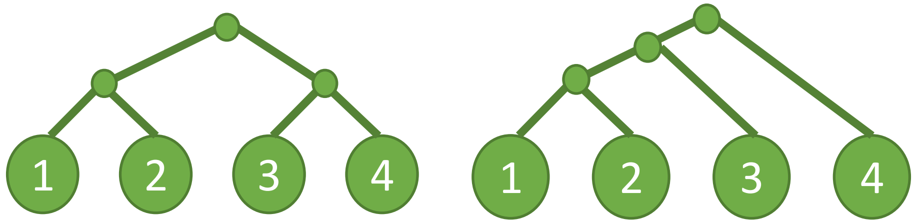
<u>Zwecks Laufzeit:</u>
Erstmal müssen wir überlegen was genau der Basisalgorithmus ist, welcher angewendet wird.
Dann überlegen wir welche Laufzeit dieser hat und dann wie oft wir diesen anwenden müssen.

Feng-Doolittle führt eigentlich immer nur paarweise Alignments durch.
→ Laufzeit für paarweises Alignment ist $O(n^2)$

Wie viele paarweise Alignments müssen wir durchführen?

Betrachten wir den obersten inneren Knoten; dort müssen wir maximal $\left(\frac{k}{2}\right)^2$ paarweise Vergleiche durchführen.

Da wir immer noch $k - 1$ innere Knoten haben
→ <u>Gesamtlaufzeit: $O(k^3 * n^2)$</u>



Revision Übung 17:

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

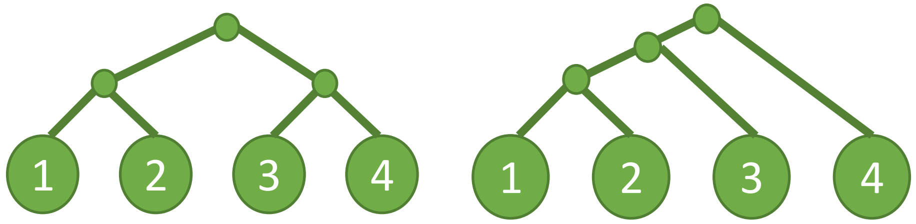
Zwecks Laufzeit:
Erstmal müssen wir überlegen was genau der Basisalgorithmus ist, welcher angewendet wird.
Dann überlegen wir welche Laufzeit dieser hat und dann wie oft wir diesen anwenden müssen.

Feng-Doolittle führt eigentlich immer nur paarweise Alignments durch.
→ Laufzeit für paarweises Alignment ist $O(n^2)$

Wie viele paarweise Alignments müssen wir durchführen?

Betrachten wir den obersten inneren Knoten; dort müssen wir maximal $\left(\frac{k}{2}\right)^2$ paarweise Vergleiche durchführen.

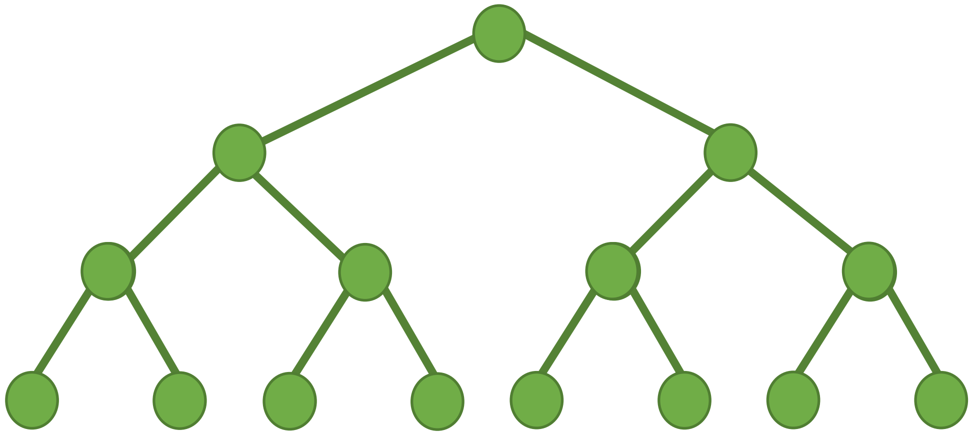
Da wir immer noch $k-1$ innere Knoten haben
→ Gesamtlaufzeit: $O(k^3 * n^2)$



Revision Übung 17:

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

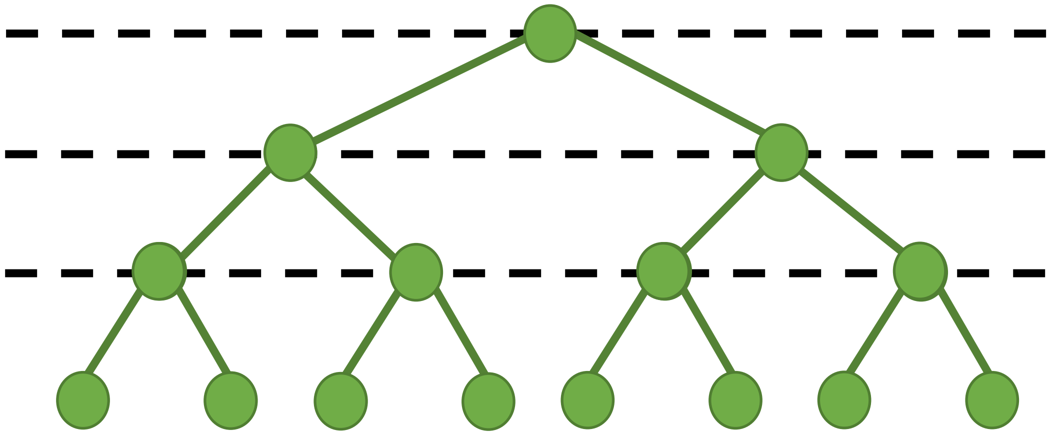
$$\left(\frac{k}{2}\right)^2$$



Revision Übung 17:

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

$$\left(\frac{k}{2}\right)^2$$



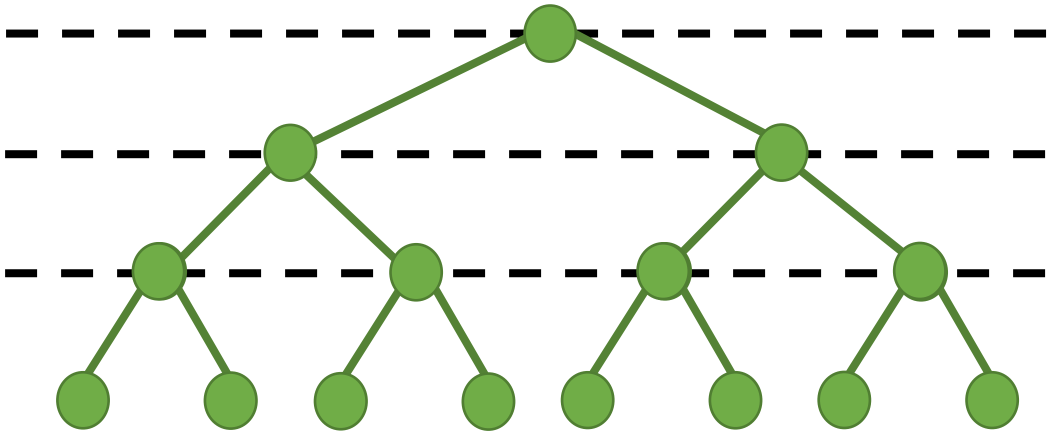
$$2 * \left(\frac{k}{4}\right)^2$$

$$4 * \left(\frac{k}{8}\right)^2$$

Revision Übung 17:

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

$$\left(\frac{k}{2}\right)^2$$



$$2 * \left(\frac{k}{4}\right)^2$$

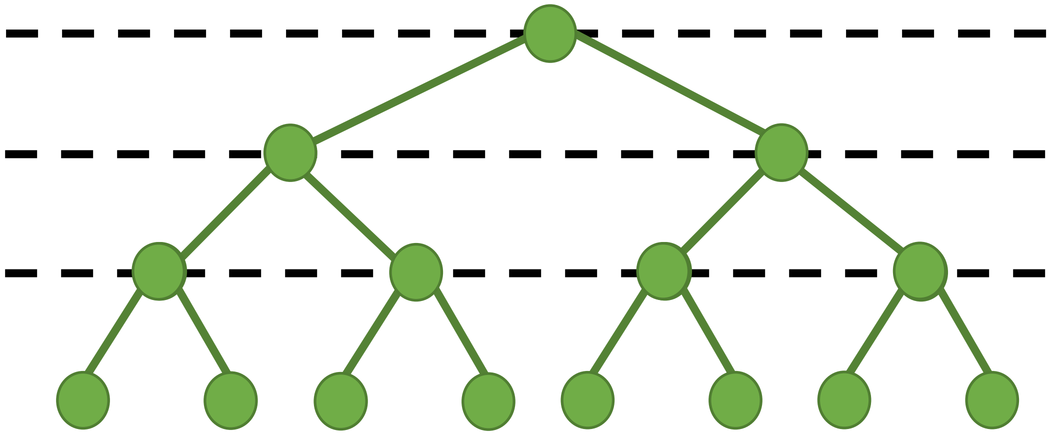
$$4 * \left(\frac{k}{8}\right)^2$$

$$\sum_{i=1}^{\log_2 k} \left(\frac{k}{2^i}\right)^2 * 2^{i-1}$$

Revision Übung 17:

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

$$\left(\frac{k}{2}\right)^2$$



$$2 * \left(\frac{k}{4}\right)^2$$

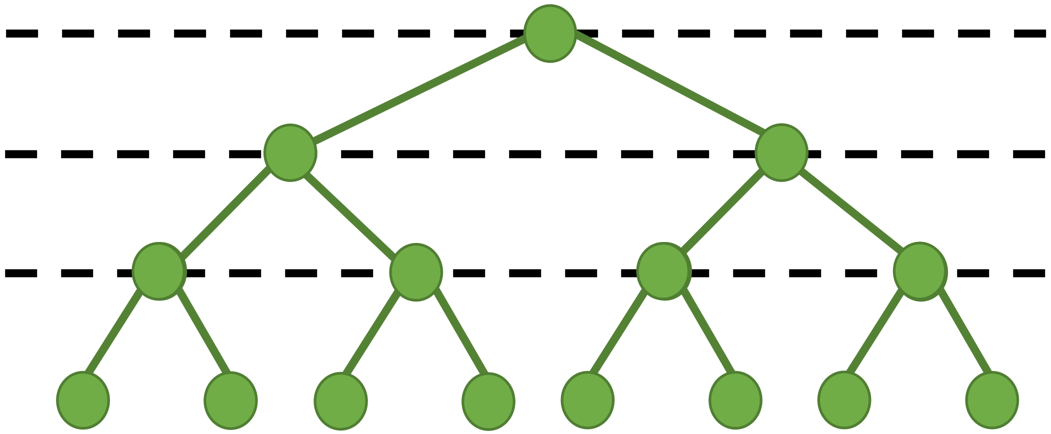
$$4 * \left(\frac{k}{8}\right)^2$$

$$\sum_{i=1}^{\log_2 k} \left(\frac{k}{2^i}\right)^2 * 2^{i-1} = O(k^2)$$

Revision Übung 17:

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum.
Annahme: Die Sequenzen dürfen sich während des Progressieven Alignemnents nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).

$$\left(\frac{k}{2}\right)^2$$



$$2 * \left(\frac{k}{4}\right)^2$$

$$4 * \left(\frac{k}{8}\right)^2$$

$$\sum_{i=1}^{\log_2 k} \left(\frac{k}{2^i}\right)^2 * 2^{i-1} = O(k^2)$$

→ <u>Gesamtlaufzeit Feng-Doolittle Verfahen: $O(k^2 * n^2)$</u>