

Der Evolution mit Hochleistungsrechnern auf der Spur

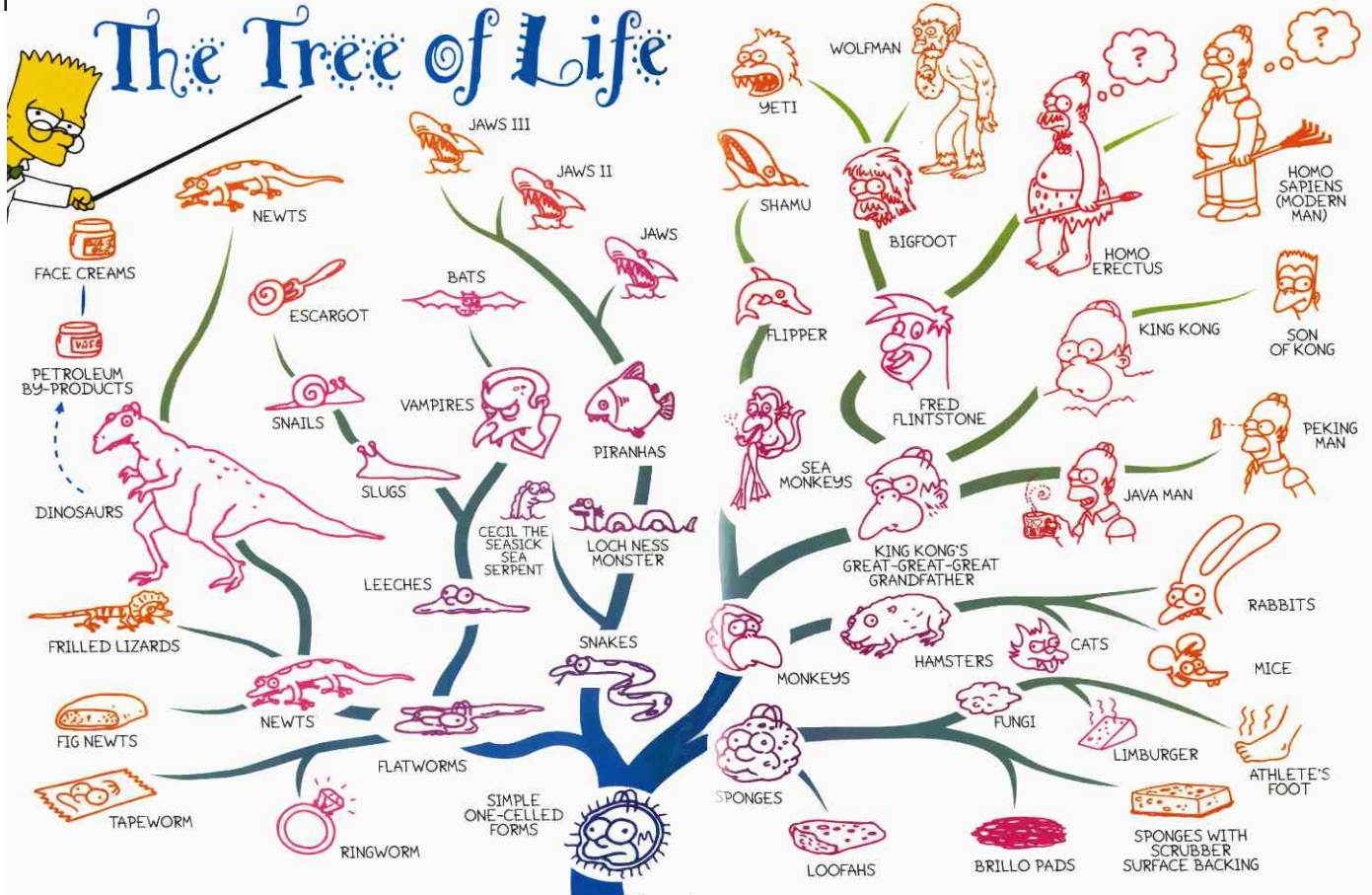


Alexandros Stamatakis
Scientific Computing Group
HITS

Fragen

- Bitte stellen Sie Fragen !

Berechnung molekularer Stammbäume



Stammbäume: Eingabe



AAGGCCCT



AAGACACCT

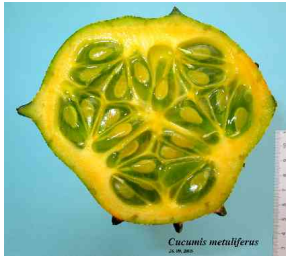


AAGCCCCCT



AAGGCCTTT

Stammbäume: Eingabe



AAGGCCCT

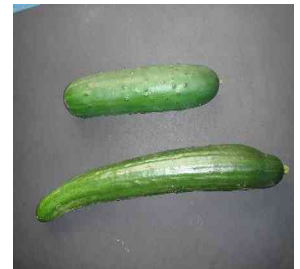
DNA (DNS)
Eingabesequenzen

AAGACACCT

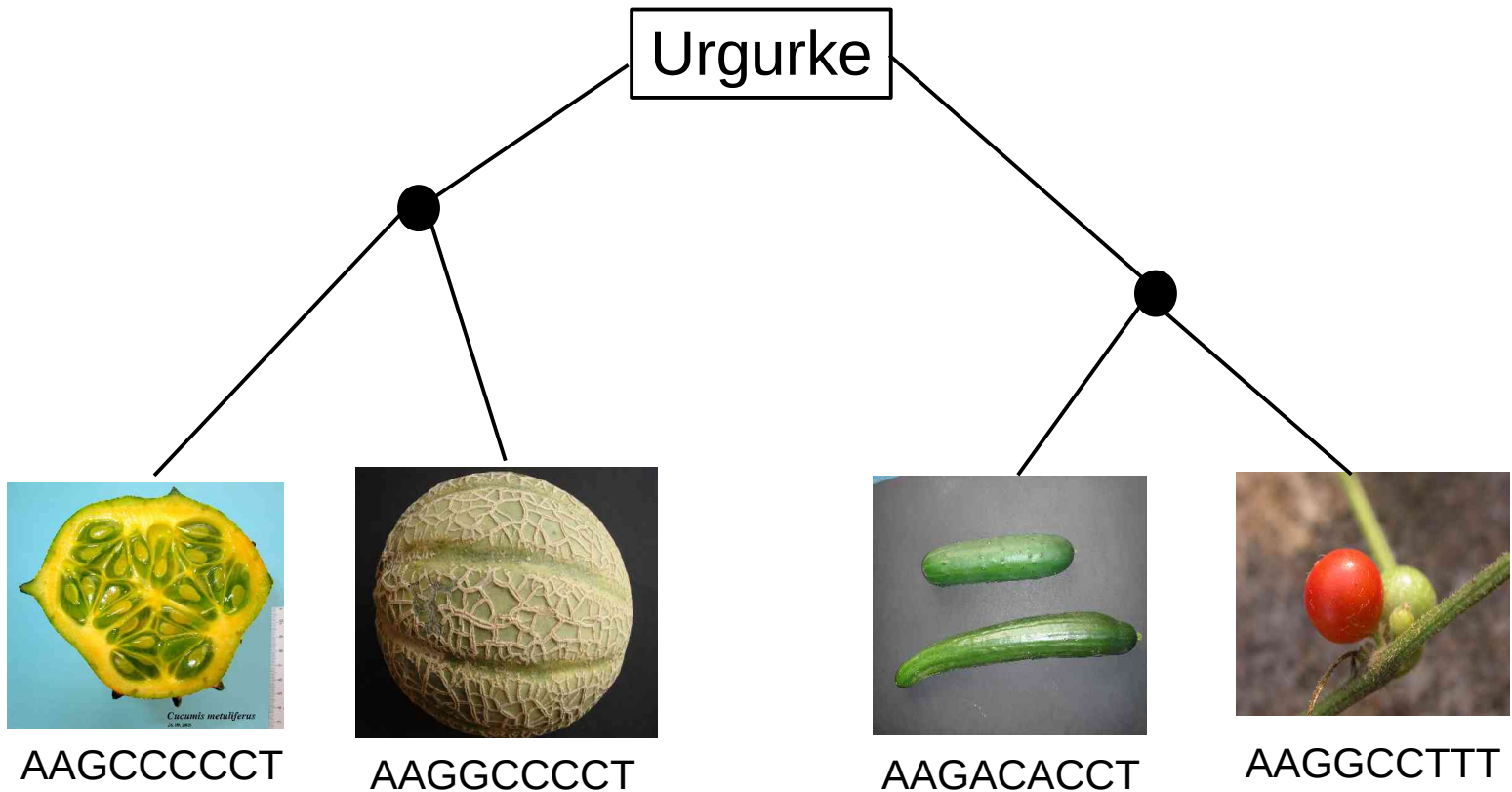


AAGCCCCCT

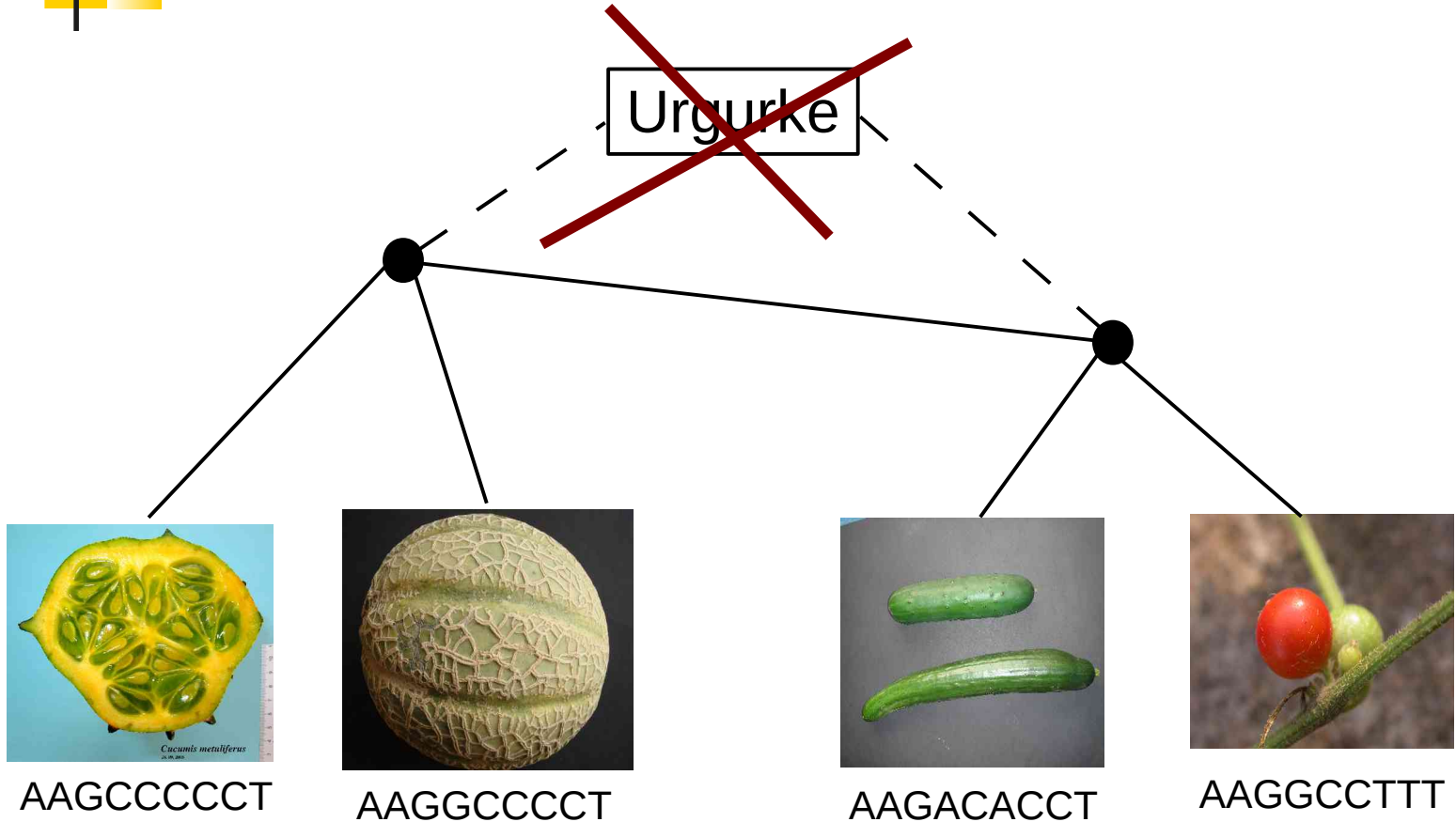
AAGGCCTTT



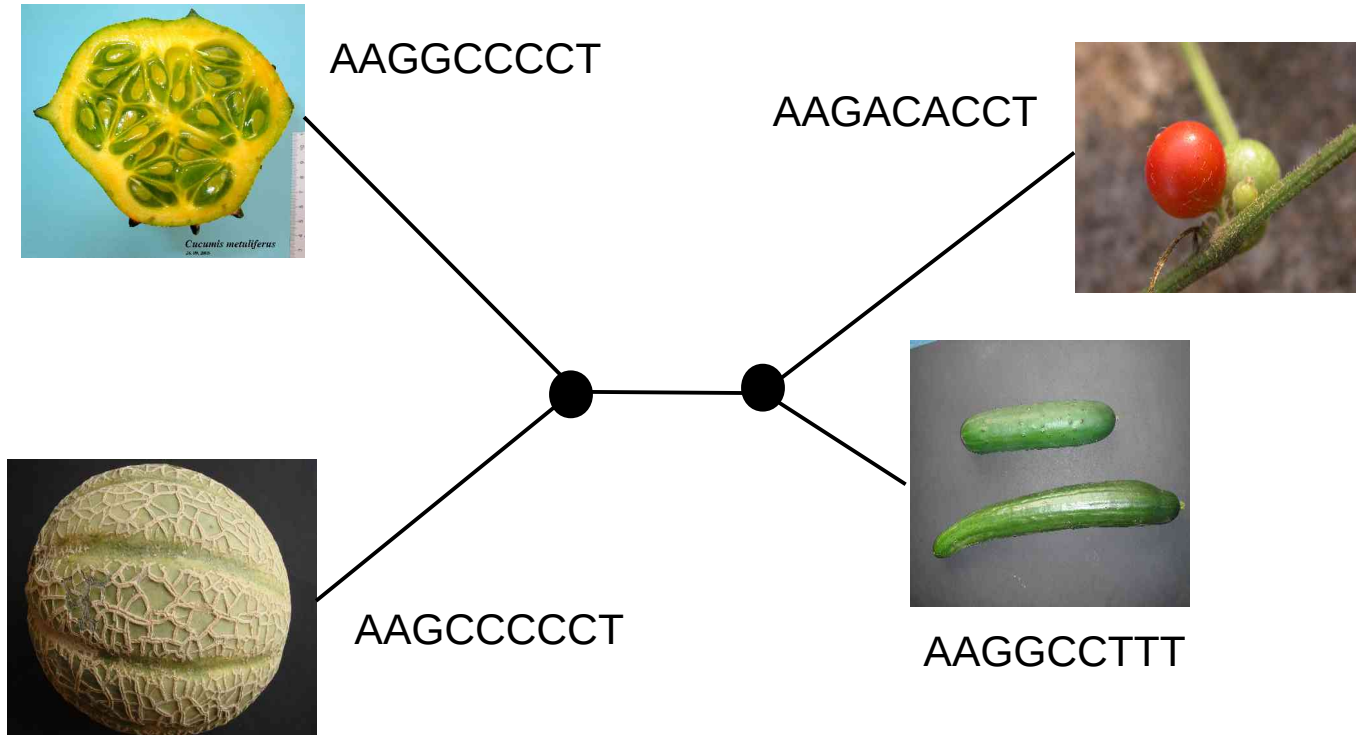
Stammbäume: Ausgabe



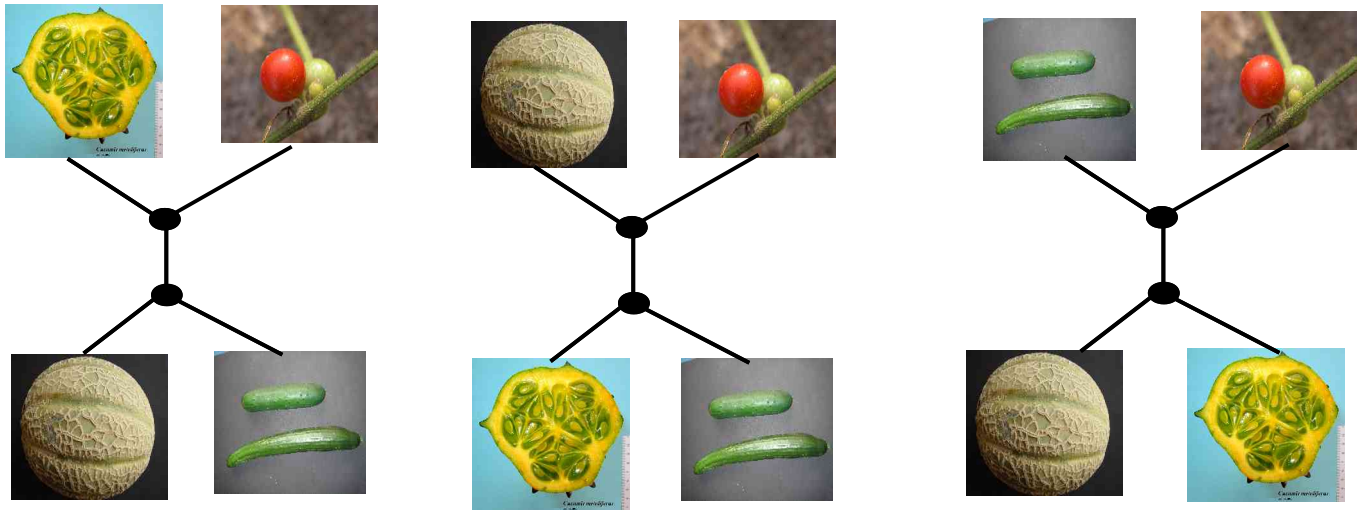
Stammbäume: Ausgabe



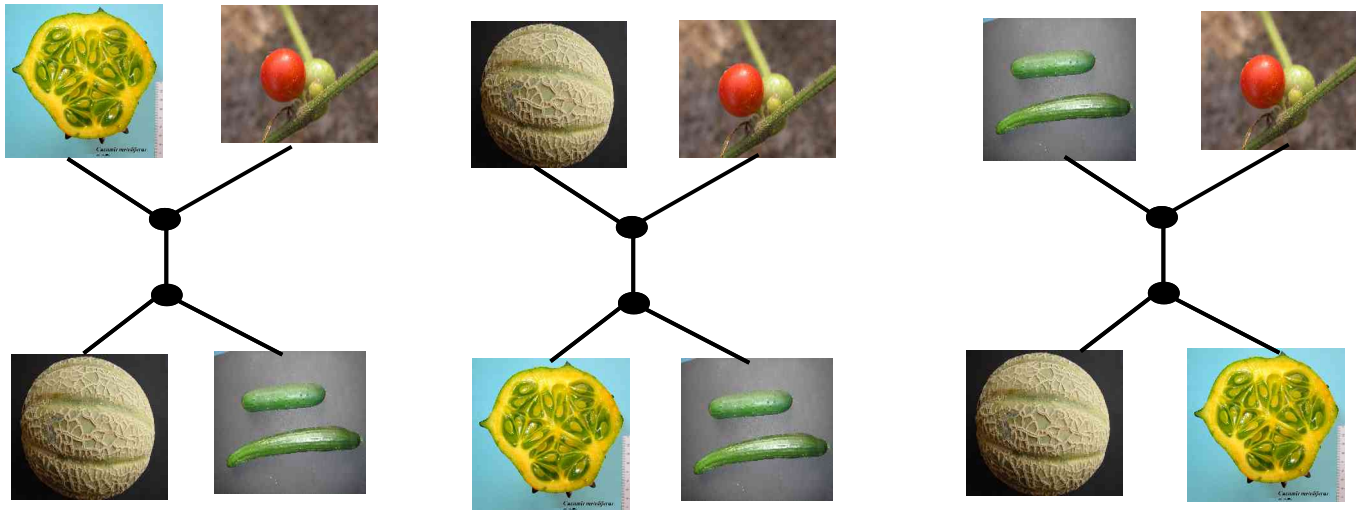
Stammbäume: Ausgabe



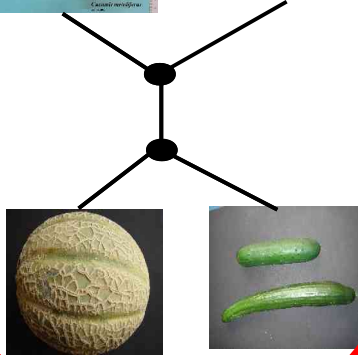
Welcher ist der beste Baum?



Wir brauchen eine Bewertungsfunktion!



Wir brauchen eine Bewertungsfunktion!



$f()$



1.0

Wir brauchen eine Bewertungsfunktion!

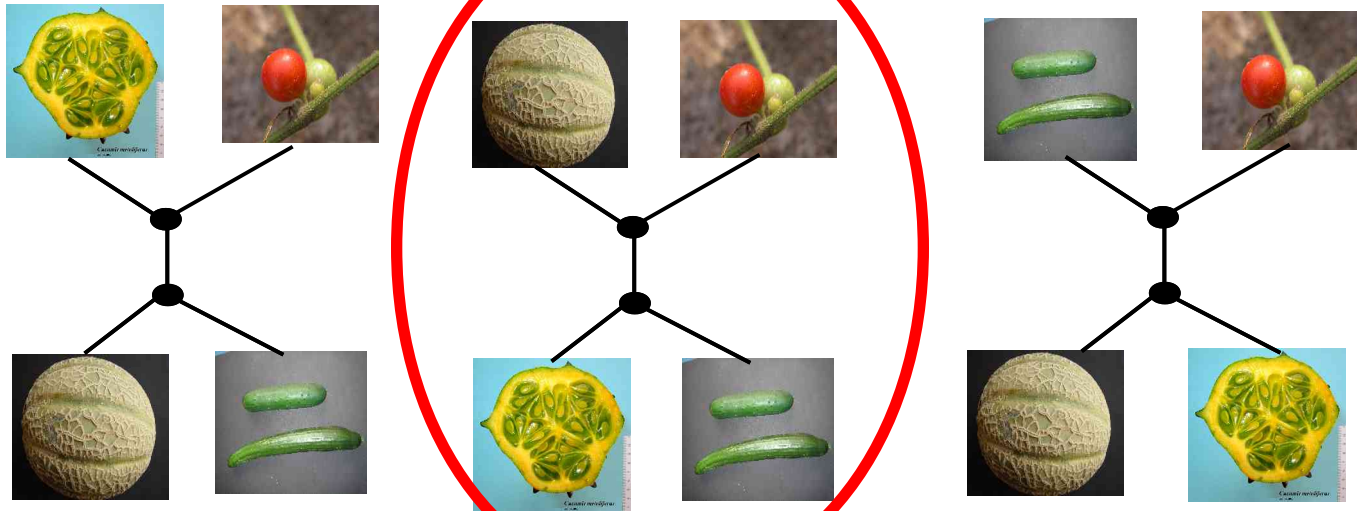
Funktion $f()$ →
mathematische
Modellbildung

Statistische Methoden:
Maximum Likelihood

$f()$

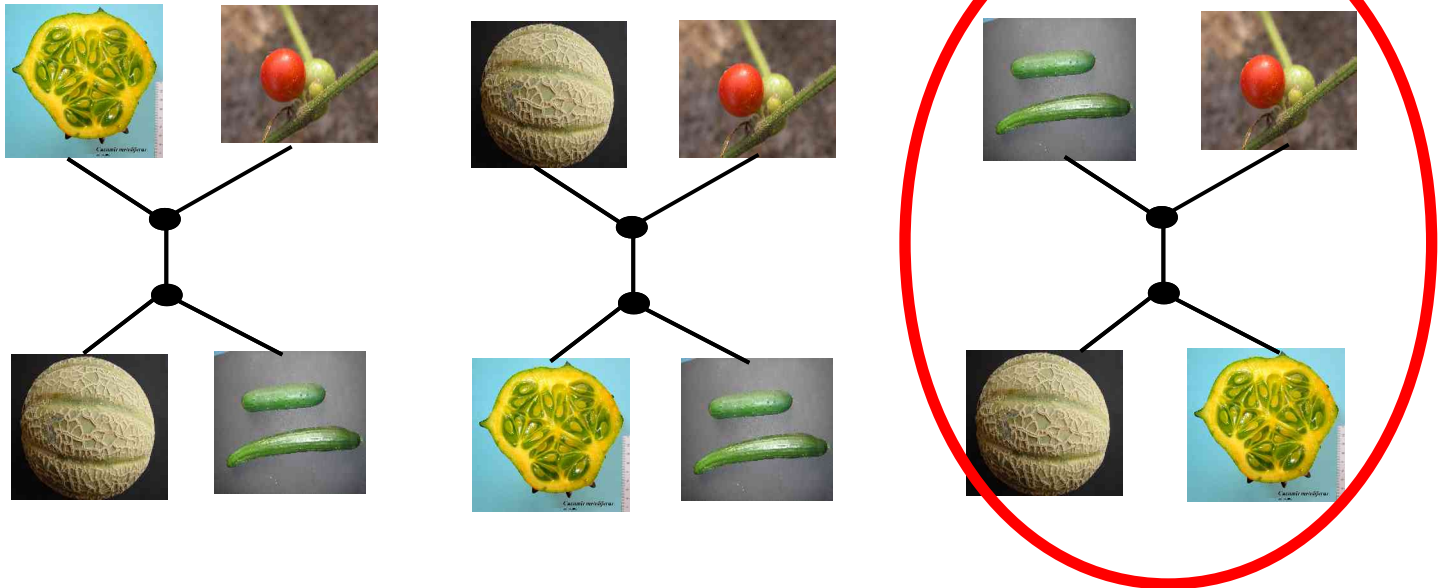
↓
1.0

Wir brauchen eine Bewertungsfunktion!



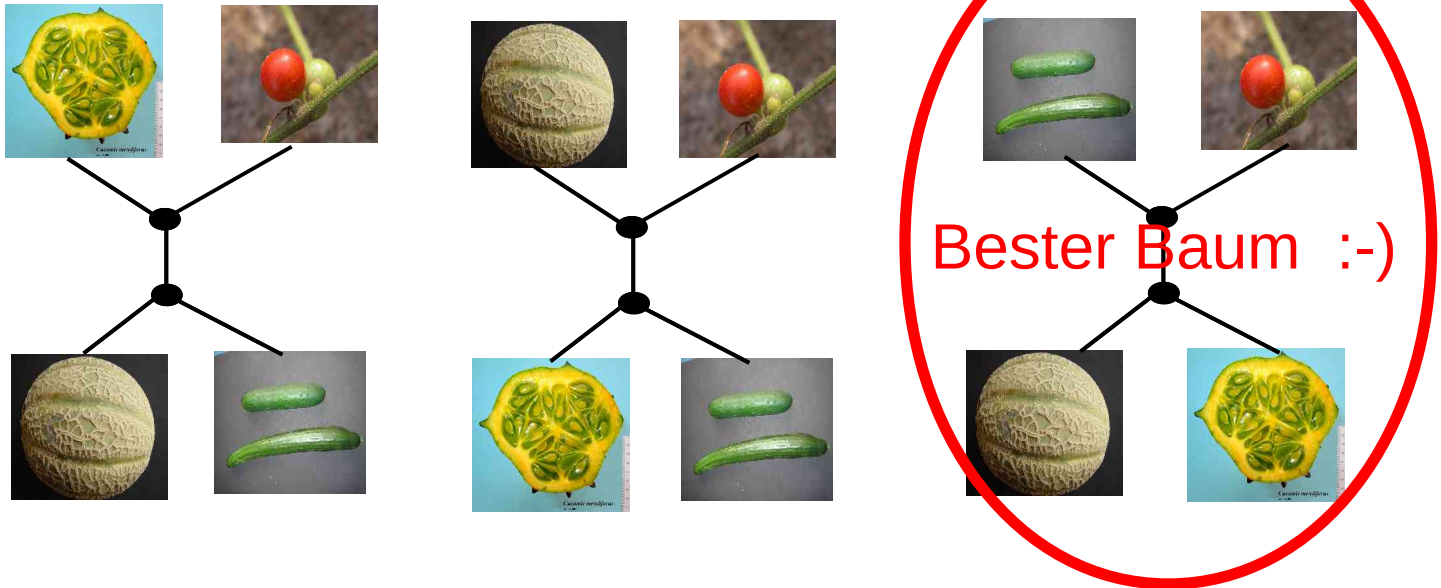
$f()$
↓
2.0

Wir brauchen eine Bewertungsfunktion!



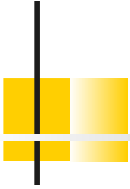
$f()$
↓
3.0

Wir brauchen eine Bewertungsfunktion!

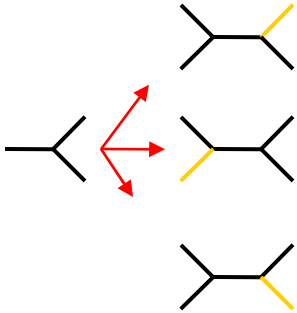
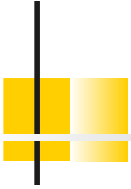


$f()$
↓
3.0

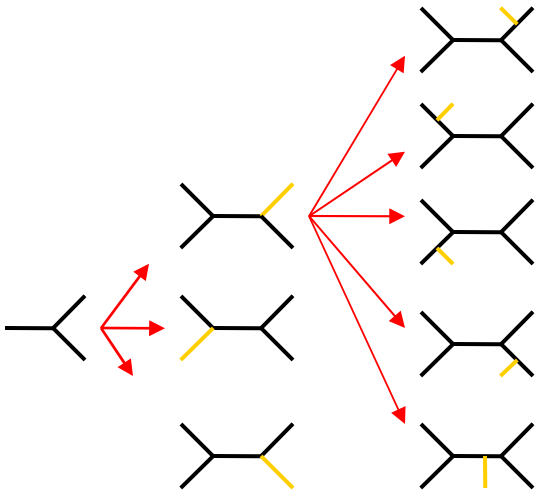
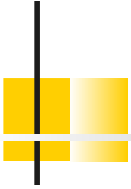
Mit dem Kopf durch die Wand



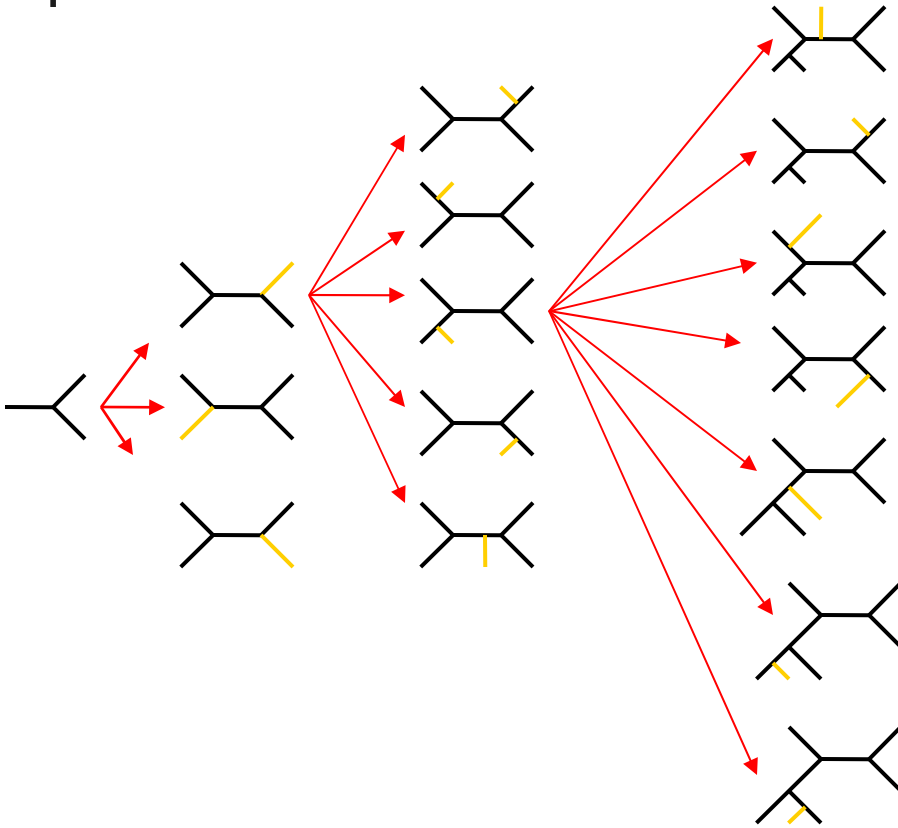
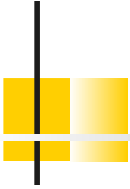
Mit dem Kopf durch die Wand



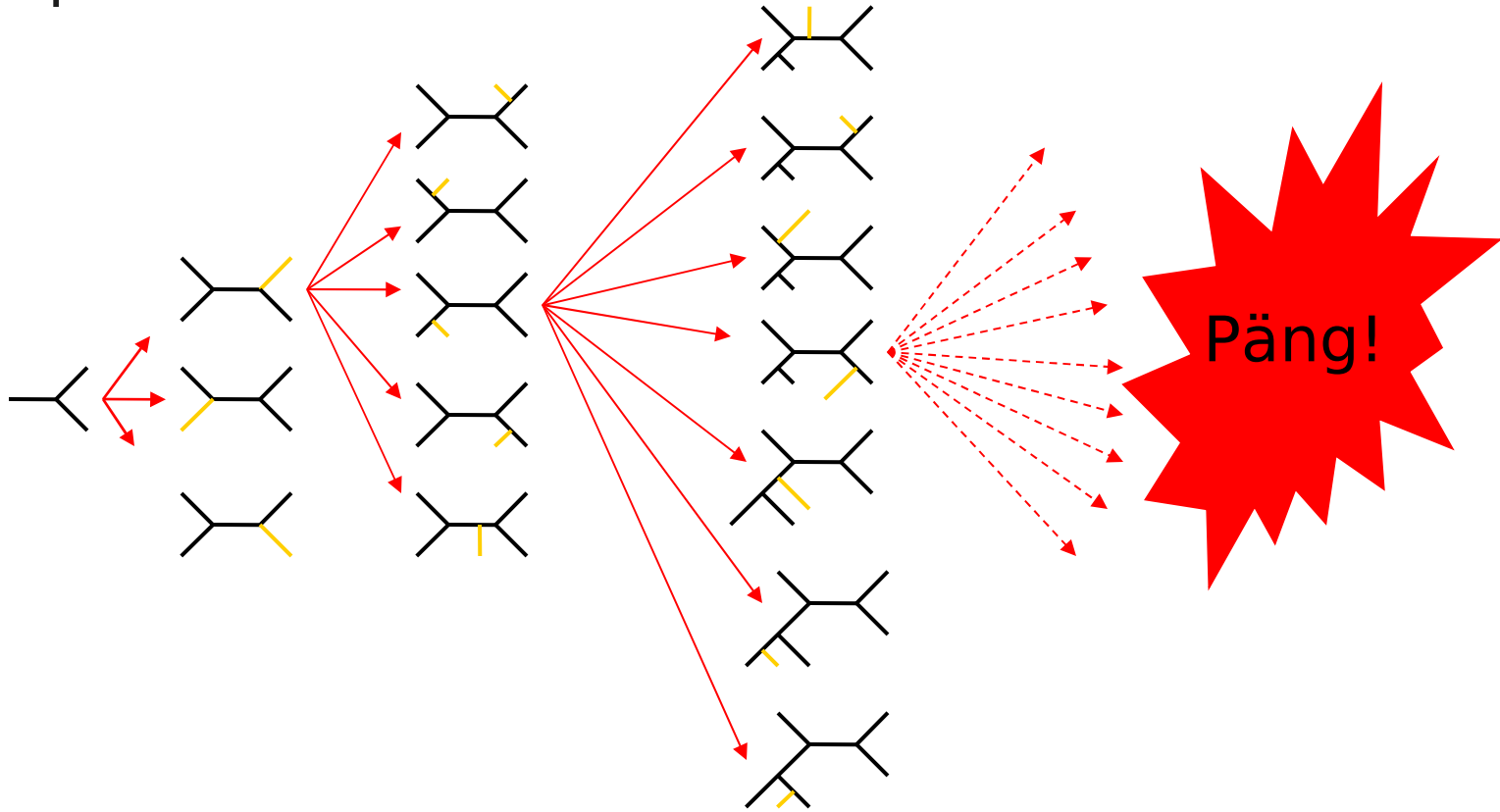
Mit dem Kopf durch die Wand



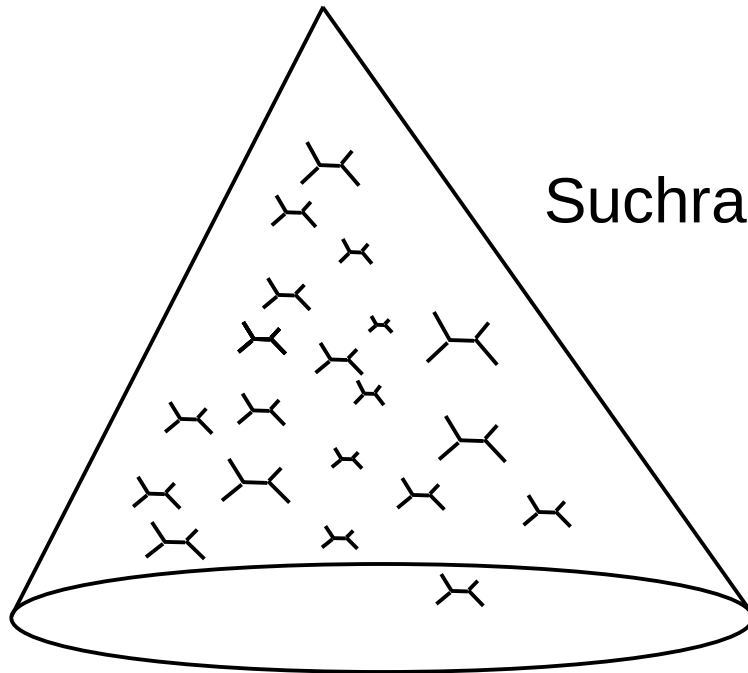
Mit dem Kopf durch die Wand



Mit dem Kopf durch die Wand



Suchstrategien



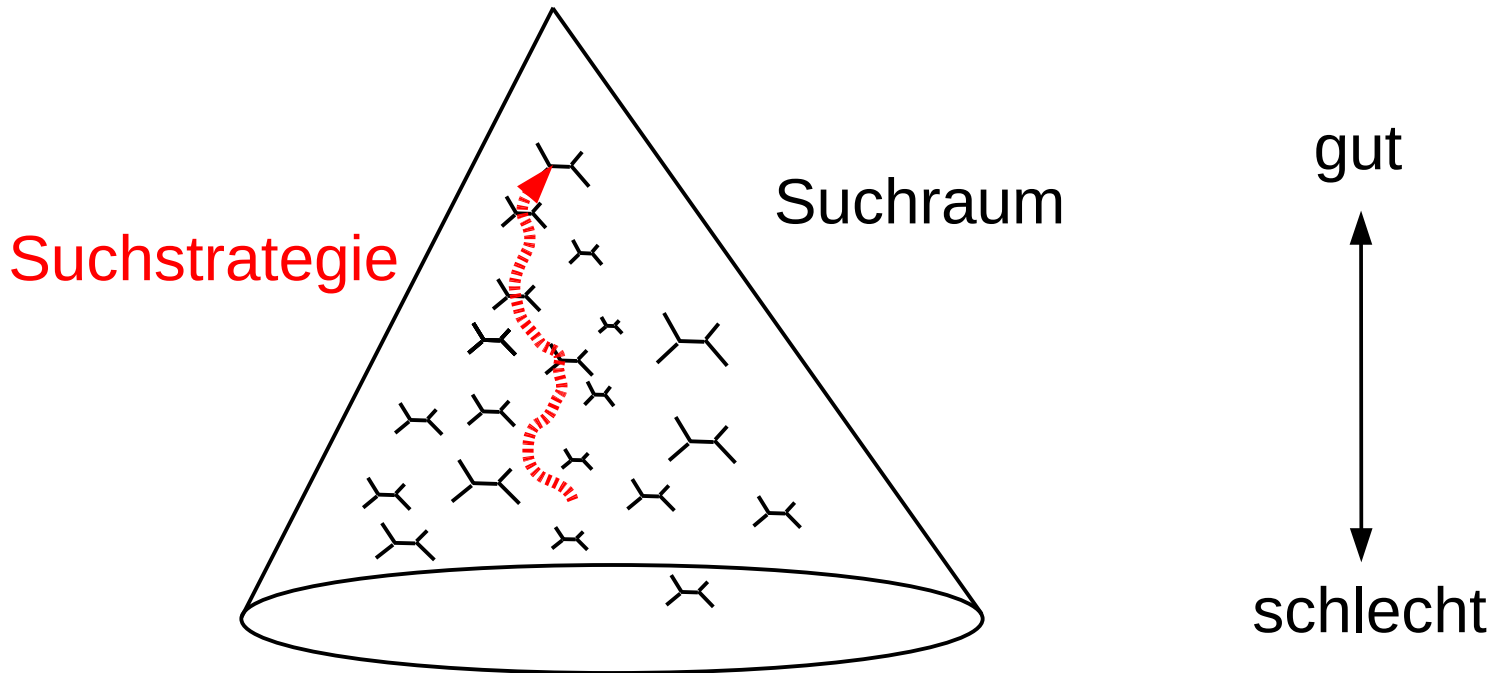
Suchraum

gut



schlecht

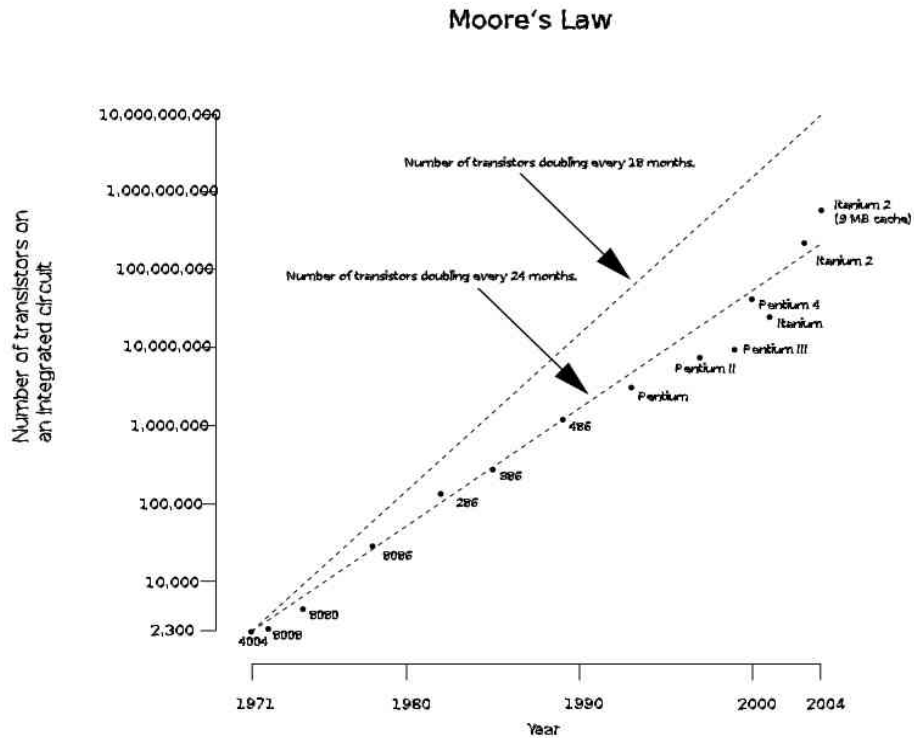
Suchstrategien



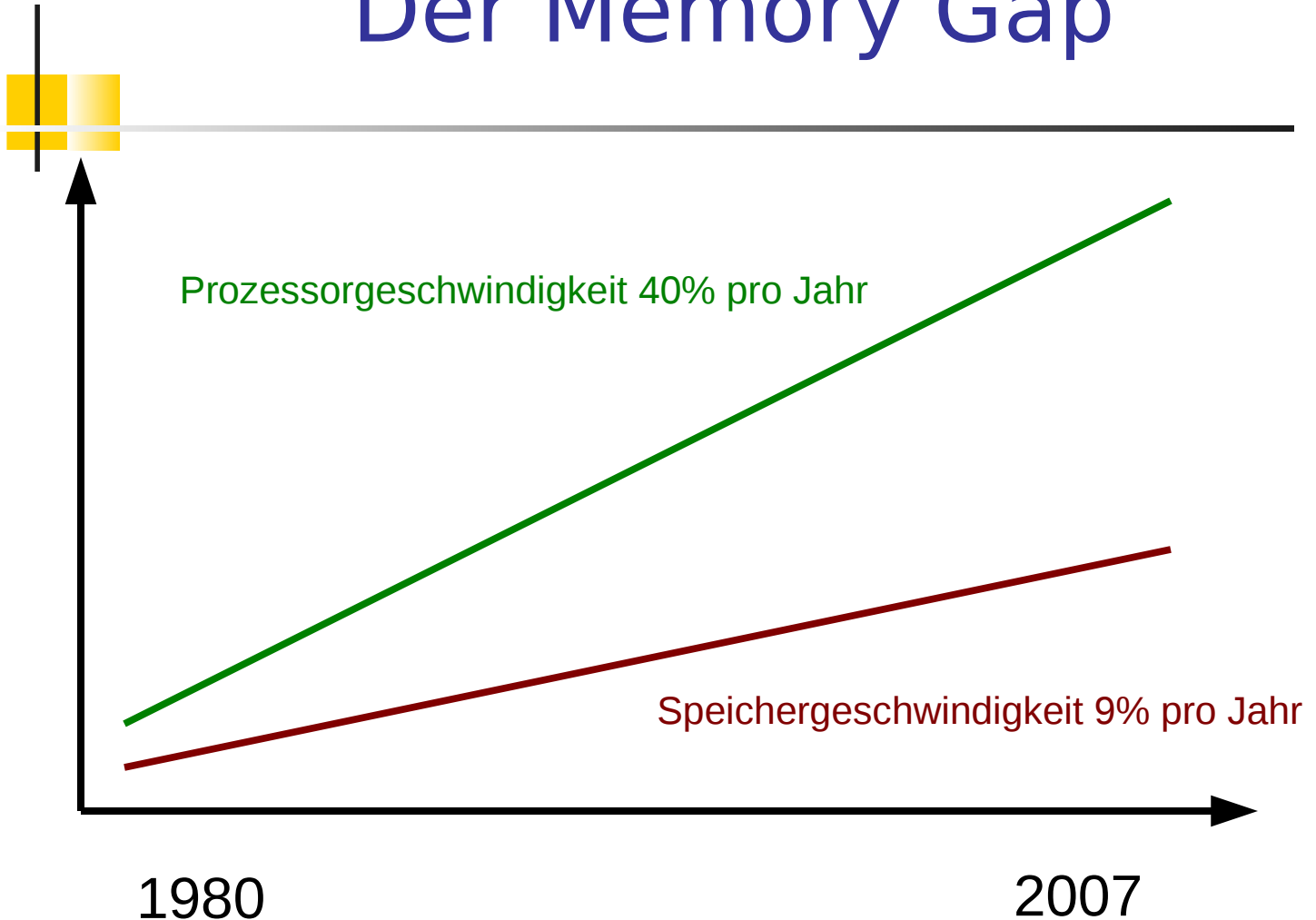
Black Box Demo



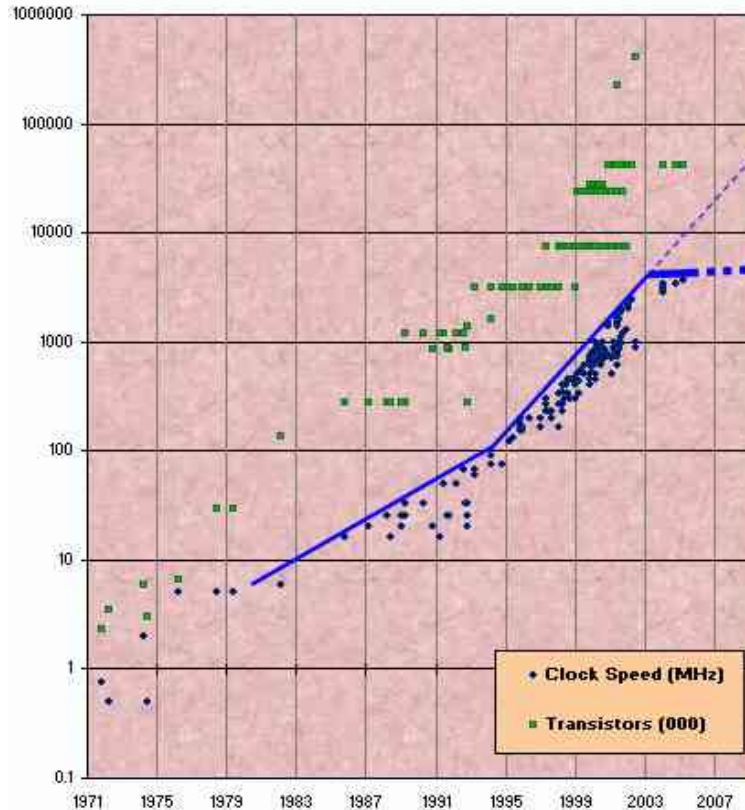
Moore's Law



Der Memory Gap



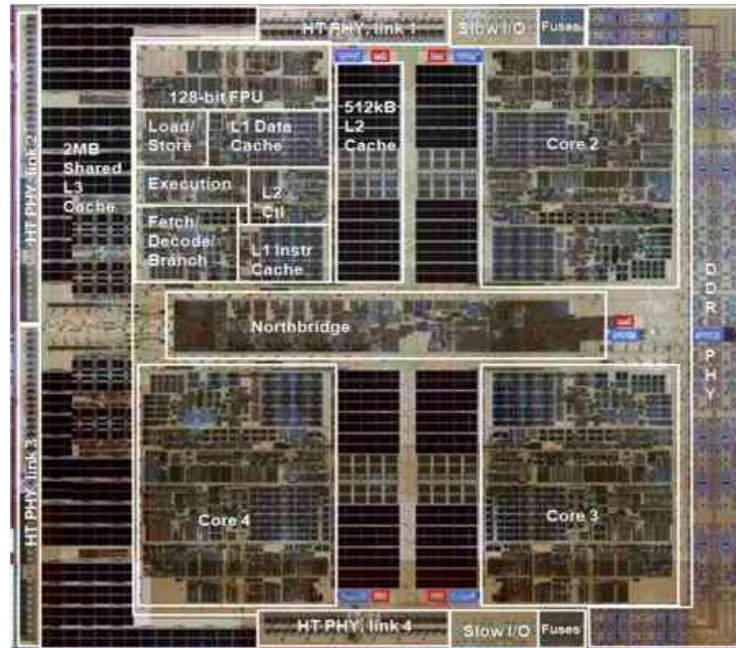
Das Freibier ist aus!



Warum kann man die Taktzahlen nicht einfach weiter erhöhen?

- Man bekommt nicht genug Daten in den Prozessor wegen Memory Gap
 - Problem mit Überhitzung: Temperatur steigt im Quadrat zur Taktfrequenz
 - Prozessoren würden bei über 5Ghz schmelzen
 - The Clock-speed Wall
- Lösung (oder auch nicht)
Paradigmenwechsel zu Mehrkern (Multi-Core) Architekturen “The free lunch is over!”

Multi-Core Architektur





Evolution der Rechnerarchitekturen

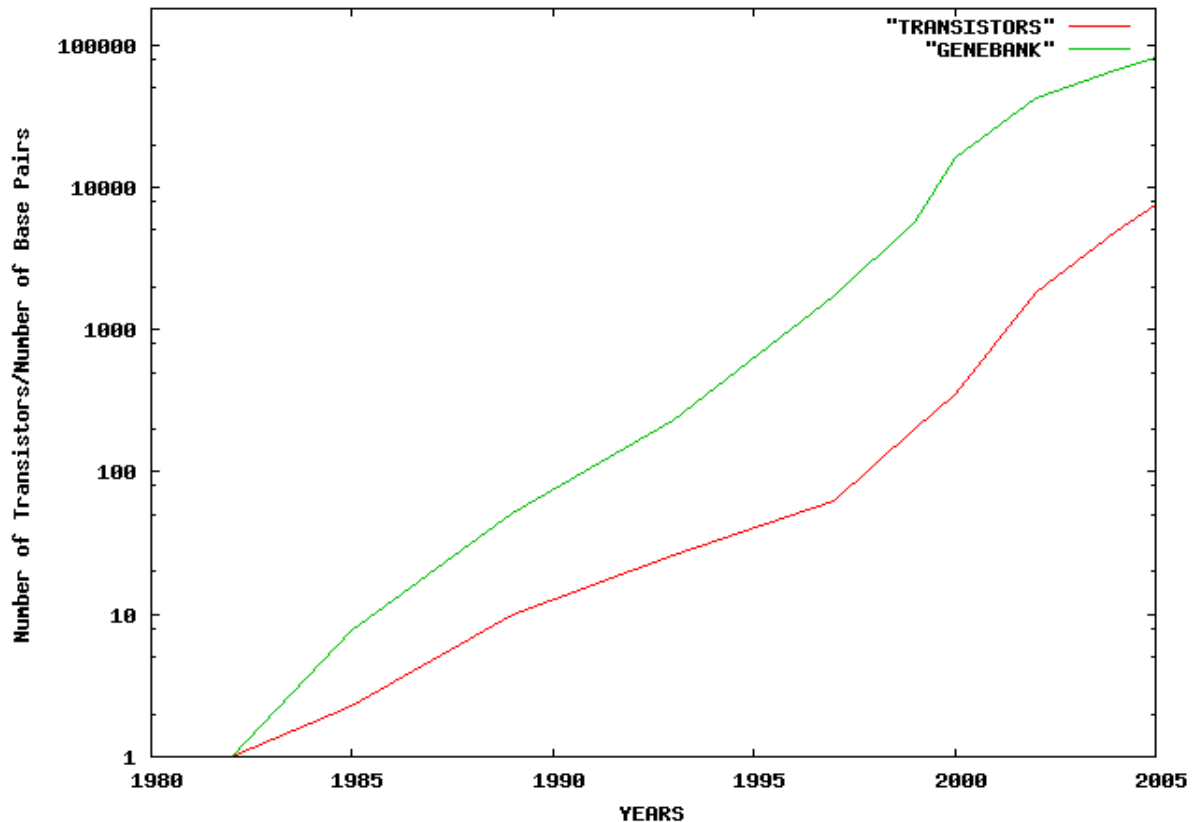
- Moore's law
- Ausserdem gibt es ein Speicherproblem
- Paradigmenwechsel zu Mehrkern-Architekturen
- Parallelisierung von Programmen kann nicht automatisiert werden



Evolution der Rechnerarchitekturen

- Moore's law
- Ausserdem gibt es ein Speicherproblem
- Paradigmenwechsel zu Mehrkern-Architekturen
- Parallelisierung von Programmen kann nicht automatisiert werden
- Neue DNA-Sequenziertechniken: Nie dagewesene Datenflut

Eine rasante Entwicklung Der "Bio-Gap"





Der “Bio-Gap”: Ein aktuelles Beispiel

- April 2008: Dunn *et al*, *Nature*: “Broad Phylogenomic Sampling Improves the Accuracy of the Animal Tree of Life”
 - Analyse konnte noch auf einem normalen (relativ günstigen) Cluster durchgeführt werden
- 2009:
 - 10 Mal grösserer Datensatz
 - 2,250,000 Prozessorstunden auf einem Supercomputer (IBM BlueGene/L)
- Gegenwärtige Analysen: 10-100 Mal grösser

Fazit



- Parallellrechnen wird zunehmend wichtiger für die Bioinformatik!
- Man muss sich zu einer neuen Sequenziermaschine einen Hochleistungsrechner dazukaufen!
- Neue Sequenziermaschine (454-Sequencer) auf Kreta → reicht für das ganze östliche Mittelmeer

Datenverteilung & Parallelisierung



ACGTT AAGGG TTTCC CCGGG



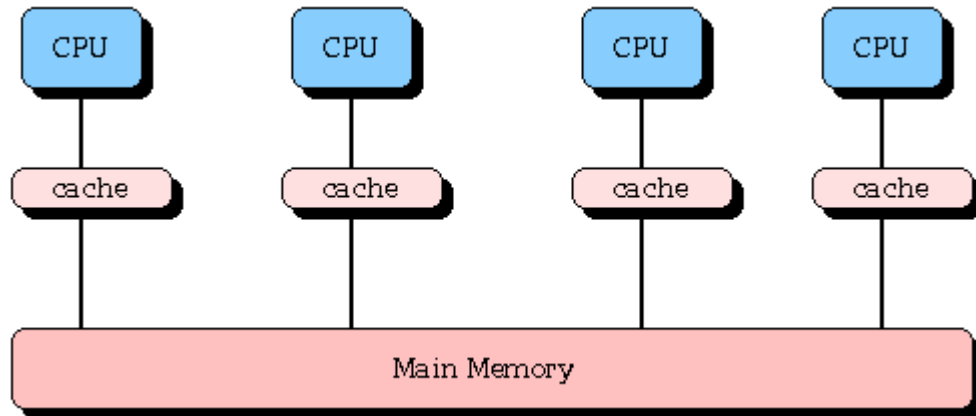
ACGTA AAAGG TTCCC CCCGG



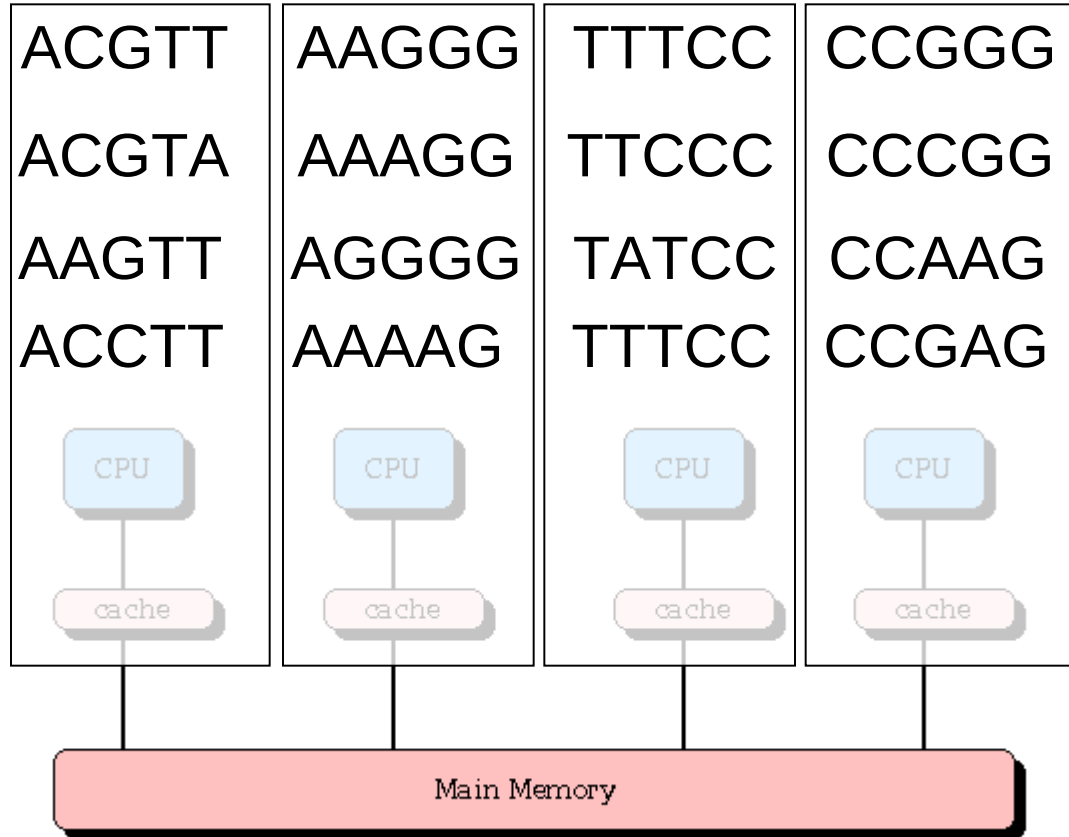
AAGTT AGGGG TATCC CCAAG



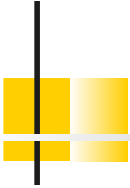
ACCTT AAAAG TTTCC CCGAG



Datenverteilung & Parallelisierung



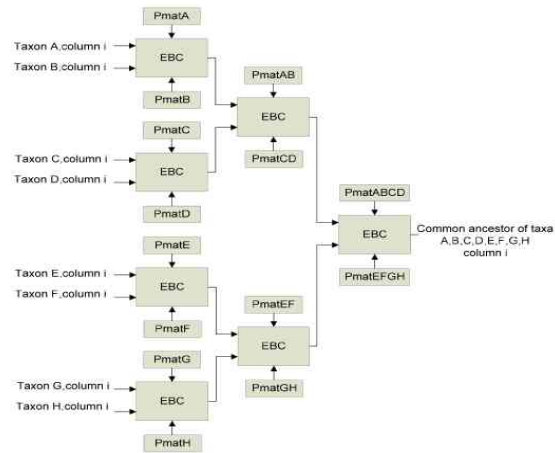
Bewertungsfunktion $f()$ als Schaltkreis



Nikos Alachiotis
Doktorand HITS

4.3.1 Το Μονοπάτι Δεδομένων – Εναλλακτική 1

Σύμφωνα με την πρώτη εναλλακτική σχεδιαστική προσέγγιση που προτείνεται, τα expanded basic cells τοποθετούνται σε δενδρική δομή όπως φαίνεται στην εικόνα 4-4.



Εικόνα 4-4 : Δενδρική Τοπολογία της επεκταμένης βασικής υπολογιστική μονάδας

Η παραπάνω σχεδιαστική επιλογή παρουσιάζει 7 expanded basic cells σε δενδρική διάταξη. Τα 4 EBCs του πρώτου επιπέδου λειτουργούν παράλληλα και επεξεργάζονται δεδομένα από 8 διαφορετικές λειτουργικές ταξινομηκές μονάδες. Η πληροφορία που προκύπτει αποτελεί είσοδο στα EBCs του δεύτερου επιπέδου και τα αποτελέσματα του

Stammbaumberechnung auf Hochleistungsrechnern



IBM BlueGene/L San Diego Supercomputer Center
LRZ HLRB-II Höchstleistungsrechner in Bayern



Michael Ott
Ehem. Doktorand

Playstation III

Nicht nur zum Spielen



Filip Blagojevic
Berkeley Labs, Kalifornien



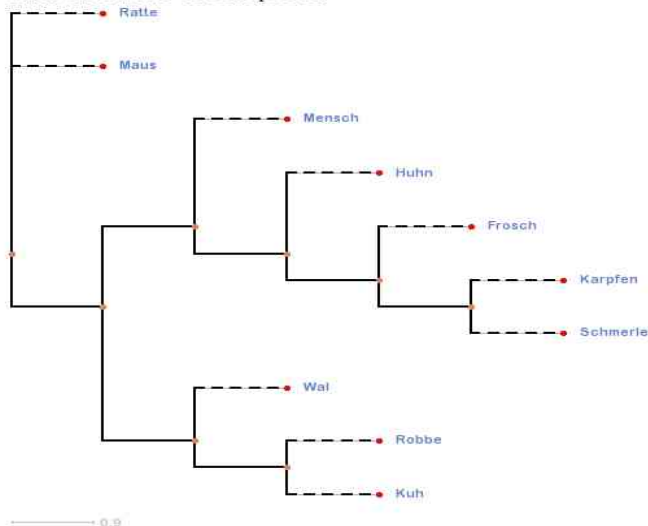
Was macht unsere Stammbaumanalyse?

Ein kleiner Stammbaum

RAxML BlackBox

Consensus tree (MRE)

Click on a node to collapse it:



When publishing results obtained via the web-server please cite:

A. Stamatakis, P. Hoover, J. Rougemont

[A Rapid Bootstrap Algorithm for the RAxML Web-Servers.](#)

Systematic Biology, 75(5): 758-771, 2008

Ausbreitung des Vogelgrippe-Virus 1996-2004





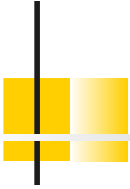
Video: Ausbreitung des Vogelgrippe-Virus 1996-2004

Zusammenfassung



- Stammbäume sind für viele Bereiche der biologischen und medizinischen Forschung wichtig
- Interdisziplinäres Forschungsfeld: Statistik, Algorithmik, Höchstleistungsrechnen, Biologie
- Die biologische Datenflut stellt uns vor neue Herausforderung

Vielen Dank für Ihre Aufmerksamkeit!



Psiloritis, Kreta, Griechenland, Januar 2006