Vorlesung

Einführung in die Praktische Informatik

Wintersemester 2015/2016

Universität Heidelberg

Organisatorisches

Vorlesung

- Dozent
  Filip Sadlo, INF 368, Raum 528, Sprechstunde: Mi 12:30–13:30 Uhr
- Übungsleiter
  TBA
  email: `ipi2015@iwr.uni-heidelberg.de`
- TutorInnen: Studierende höherer Semester
- Webseite zur Vorlesung
  `www.iwr.uni-heidelberg.de/groups/viscomp/teaching/2015-16/ipi/`
  Informationen, Unterlagen, Übungsblätter
- Skript
  Basierend auf Bastian (2003,2011,2014), Neuss (2006)

Übungen

- Sinn der Übungsgruppen
  Vertiefen Stoff der Vorlesung
  Jede Woche wird ein Blatt mit Aufgaben ausgegeben
  Besprechen der Aufgaben in den Übungsgruppen
  Fragen zur Vorlesung und den Übungen stellen

- Derzeit 20 Übungsgruppen, Abgabe in Gruppen je 2–3 Teilnehmer
  (keine Einzel-Abgaben!)

- Anmeldung über MÜSLI
  `https://www.mathi.uni-heidelberg.de/muesli/lecture/view/538`
  Anmeldungszeitraum: bis Do., 15.10., 23:59 Uhr
  Einteilung am Freitag Vormittag (aktuelle Einteilung gilt nicht!)
  System arbeitet mit Präferenzen und nicht first come first served
  Ergebnis der Einteilung: In MÜSLI einloggen!

Ablauf der Übungen

- Ausgabe der Übungsblätter
  Donnerstags, 16 Uhr auf der Webseite der Vorlesung

- Abgabe der Lösungen
  Donnerstags, 14 Uhr ct (vor der Vorlesung)
  Abgabe in Zettelkästen am INF 288, in der Ecke zwischen HS 2 und Seifertraum

- Erstes Blatt
  Ausgabe: 15.10.15, 16 Uhr (diese Woche!)
  Abgabe: 22.10.15, 14 Uhr

- Beginn der Übungsgruppen ab Montag, 19.10.
  Kennenlernen, Fragen zu den Aufgaben und der Vorlesung

Übungszeiten

	Mo	Di	Mi	Do	Fr
09 – 11		LA1 Vorl.	ANA1 Vorl.	LA1 Vorl.	348/SR 15 368/532
11 – 13	TI Vorl.	288/HS 5 368/248 Mathe f. Inf. Vorl.	350/U013 368/248 348/SR 13 TI Vorl.		ANA1 Vorl.
14 – 16	368/432 325/SR 24	Einf. Prak. Inf. Vorl.		Einf. Prak. Inf. Vorl.	
16 – 18	Pool	368/248 325/SR 24 368/532	368/532 348/SR 13 350/U014	368/220 348/SR 13 Mathe f. Inf. Vorl.	
18 – 20		350/U013 350/U014		350/U014	

Leistungsnachweis

- Erfolgreiche Übungsteilnahme
  mindestens 50% der Punkte aus den Übungsaufgaben
  sind Voraussetzung zur Teilnahme an der Klausur!
  Präsentation einer Lösung erwünscht

- Klausur
  voraussichtlich Donnerstag, 4. Februar 2016, 14-17 Uhr

- Wer bereits die erfolgreiche Teilnahme an den Übungen zu dieser Vorlesung vom letzten Jahr nachweisen kann und zur Prüfung angetreten ist, ist zur Klausur zugelassen.

- Für BA Informatik, LA Informatik ist diese Klausur die Orientierungsprüfung

Unterschiedliche Vorkenntnisse

Angebote für Anfänger

- Nächste Woche: „Grundlagen der Bedienung von UNIX-Systemen“
  Mo 19.10. 16-18 Uhr INF 350, CIP-Pool im UG U011/012
  Di 20.10. 18-20 Uhr INF 350, CIP-Pool im UG U011/012
  Anmeldung gleich!

- Betreutes Programmieren
  Mo 16-18 OMZ INF 350 U011/12 (50 Plätze)

Ich freue mich über Fragen! Es gibt keine dummen Fragen!

Angebot für Fortgeschrittene

- Alternativer Punkteerwerb: Ersetzen ausgewählter Übungsaufgaben durch Kleinprojekte
- Details werden auf Übungsblatt/Webseite erklärt

Programmierkurs

- Einführung in Python unabhängig von der Vorlesung
- Block-Veranstaltung: 04.04.–15.04.2016
- Pflichtveranstaltung für BA Informatik, LA Informatik + Mathe in Semester 1, LA Informatik + X in Semester 3, freiwillige Teilnahme möglich

Praktisches Üben

- Programmieren ist wesentlicher (nicht alleiniger) Inhalt der Vorlesung

- Beim Programmieren gilt: Übung macht den Meister ! Programmieren ist eine Kunst. Eines der berühmtesten Bücher der Informatik von Donald E. Knuth heißt „The Art of Computer Programming“
  Nutzen Sie alle gebotenen Möglichkeiten zum Üben!

- In der Vorlesung/Übung benutzen wir eine UNIX-Programmierumgebung. Sie sollten Zugang zu so einem System haben um die Übungen durchführen zu können. Geeignet sind LINUX, Mac oder ein Windows-System mit WUBI
  `http://wiki.ubuntuusers.de/Wubi`

- Falls Sie Schwierigkeiten haben, melden Sie sich bei ihrem Tutor

Dozent

- Filip Sadlo
  2003 Diplom Informatik (ETH Zürich)
  2010 Promotion Informatik (ETH Zürich)
  2014 Vertretungsprofessur Wissenschaftliche Visualisierung (IWR, U Heidelberg)
  2015 AG Visual Computing (IWR, U Heidelberg)

- Arbeitsgebiete
  Wissenschaftliche Visualisierung
  Volumenrendering, Visualisierung im Kontext Simulation und Physik
  Anwendungen: Strömungsmechanik, Astrophysik, etc.

Motivation

Was ist Informatik

Wissenschaft von der systematischen Verarbeitung von Information, besonders der automatischen Verarbeitung mit Hilfe von Digitalrechnern

Wikipedia, Duden Informatik

Computer science is no more about computers than astronomy is about telescopes, biology is about microscopes or chemistry is about beakers and test tubes.

Michael R. Fellows and Ian Parberry, Computing Research News, January 1993

Inhalt der Vorlesung

- Grundlegende Konzepte der Informatik kennenlernen
  z. B. Algorithmenbegriff, Komplexität, Abstraktion, . . .

- Algorithmisches Denken schulen
  Problem $\rightarrow$ Algorithmus $\rightarrow$ Programm

- Programmieren im Kleinen
  verschiedene Programmierstile (funktional, prozedural, objektorientiert, generisch), Erlernen der Programmiersprache C++
  Aber: Vorbereitung für Programmieren im Großen!

- Grundlegende Algorithmen und Datenstrukturen
  Suchen, sortieren, . . .
  Listen, Felder, Heaps, Stacks, Graphen, Bäume, . . .

Informatik als Wissenschaft

- Wortschöpfung aus „Information“ und „Automatique“ erstmals benutzt von Philippe Dreyfus (1962, laut Wikipedia).

- Grundlagen waren
  - Theorie der Berechenbarkeit (Turing, Church, 1937)
  - Entwicklung elektromechanischer/elektronischer Rechenmaschinen (Z3, 1941, ENIAC, 1946)
  - entsprechende Anwendungen (Kryptographie, ballistische Berechnungen, Differentialgleichungen lösen)

- Erster deutscher Informatikstudiengang WS 1968/69 in Karlsruhe

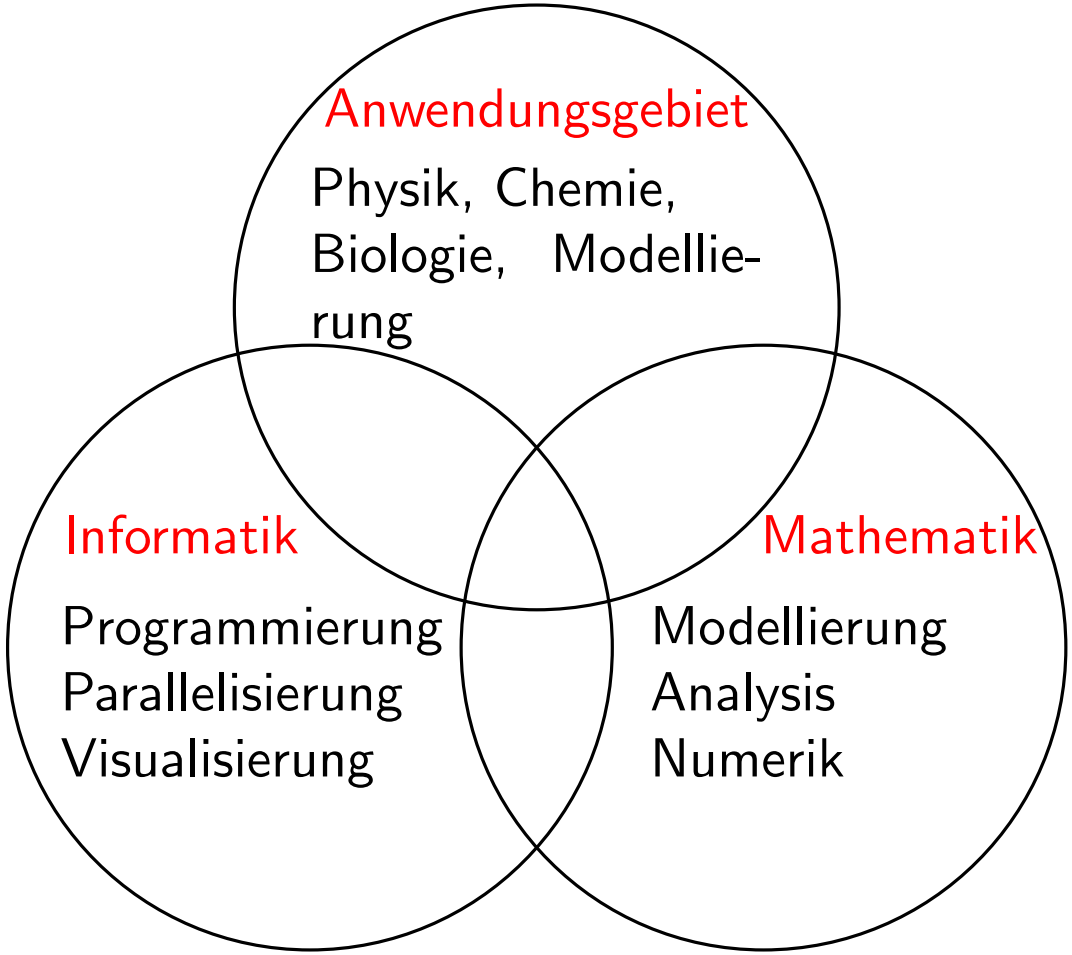
Teilgebiete der Informatik

- Theoretische Informatik
  Logik und Berechenbarkeit, Automatentheorie und formale Sprachen, Semantik, Komplexitätstheorie

- Technische Informatik
  Elektrotechnische Grundlagen, Architektur von Rechenanlagen, Chipentwurf, Netzwerkkomponenten, Fehlertoleranz, . . .

- Praktische und Angewandte Informatik
  Betriebssysteme, Softwareengineering, Datenbanken, Programmiersprachen, Visualisierung, Mensch-Maschine-Interaktion
  Anwendungen, z. B. Wissenschaftliches Rechnen

Wissenschaftliches Rechnen



Ableitung

f eine Funktion in einer Variablen:

$$f : \mathbb{R} \to \mathbb{R}$$

Ableitung:

$$\frac{df}{dx}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Partielle Ableitung

f eine Funktion in mehreren Variablen (x, y, z, t):

$$f : \mathbb{R}^3 \to \mathbb{R}$$

Partielle Ableitung bezüglich der Variablen x:

$$\frac{\partial f}{\partial x}(x, y, z) = \lim_{h \to 0} \frac{f(x + h, y, z) - f(x, y, z)}{h}$$

Differentialoperatoren

Gradient:

$$\nabla f(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y,z) \\ \frac{\partial f}{\partial y}(x,y,z) \\ \frac{\partial f}{\partial z}(x,y,z) \end{pmatrix}$$

Divergenz einer vektorw. Funktion $f(x,y,z) = (f_x(x,y,z), f_y(x,y,z), f_z(x,y,z))^T$:

$$\nabla \cdot \vec{f}(x,y,z) = \frac{\partial f_x}{\partial x}(x,y,z) + \frac{\partial f_y}{\partial y}(x,y,z) + \frac{\partial f_z}{\partial z}(x,y,z)$$

Laplace einer skalaren Funktion:

$$\Delta f(x,y,z) = \frac{\partial^2 f}{\partial x^2}(x,y,z) + \frac{\partial^2 f}{\partial y^2}(x,y,z) + \frac{\partial^2 f}{\partial y^2}(x,y,z)$$

Differentialgleichungen

Gewöhnliche Differentialgleichung:

$$\frac{df}{dt}(t) = g(t, f(t)) \qquad t \in [a, b], \qquad f(a) = g_0$$

Partielle Differentialgleichung:

$$\frac{\partial^2 f}{\partial x^2}(x, y, z) + \frac{\partial^2 f}{\partial y^2}(x, y, z) + \frac{\partial^2 f}{\partial y^2}(x, y, z) = g(x, y, z) \qquad (x, y, z) \in \Omega \subset \mathbb{R}^3$$

$$f(x, y, z) = \gamma(x, y, z) \qquad (x, y, z) \in \partial\Omega$$

Sternentstehung (Strömungsmechanik)

Cone nebula from `http://www.spacetelescope.org/images/heic0206c/`

Euler-Gleichungen

Ein Modell zur Sternentstehung stellen die Euler-Gleichungen (Leonhard Euler, 1707-1783) der Gasdynamik mit Gravitation dar. Diese sind ein nichtlineares System partieller Differentialgleichungen:

$$\partial_t \rho + \nabla \cdot (\rho v) = 0 \qquad \text{(Massenerhaltung)}$$

$$\partial_t (\rho v) + \nabla \cdot (\rho v v^T + pI) = -\rho \nabla \Psi \qquad \text{(Impulserhaltung)}$$

$$\partial_t e + \nabla \cdot ((e+p)v) = -\rho \nabla \Psi \cdot v \qquad \text{(Energieerhaltung)}$$

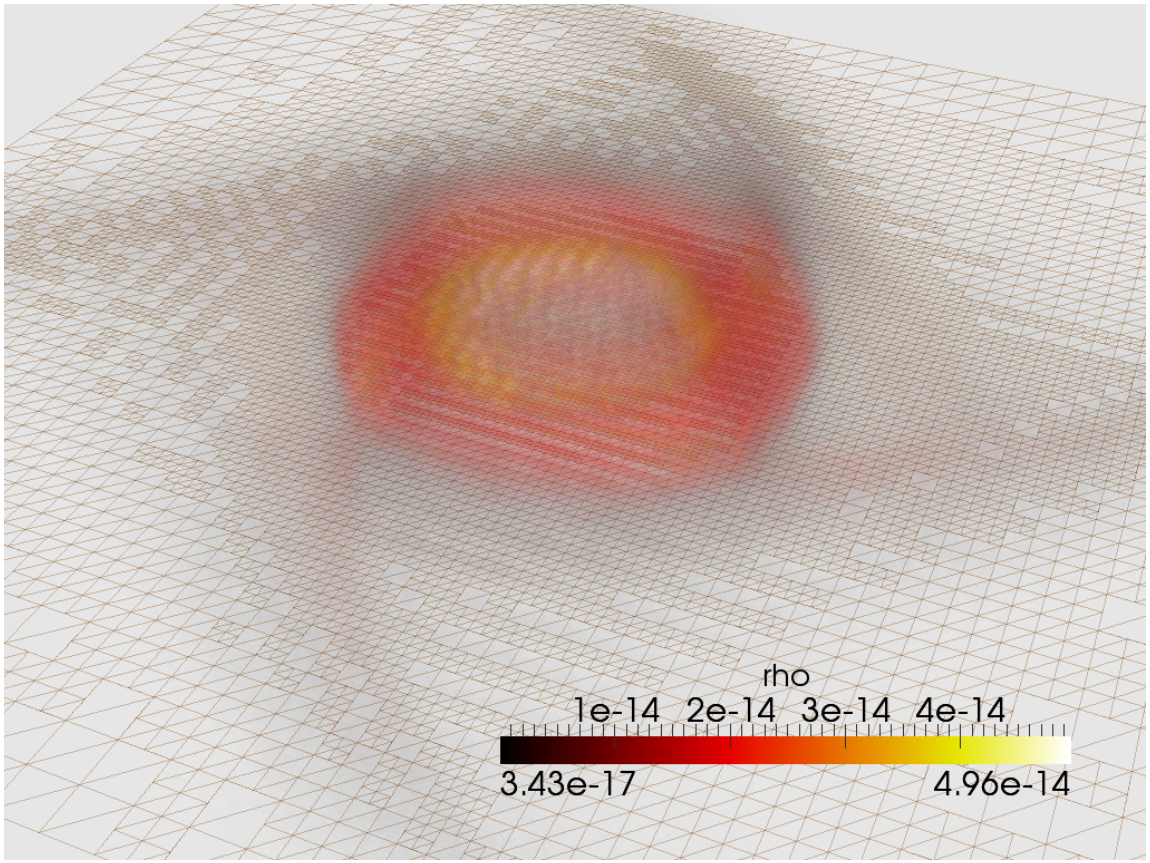
$$\Delta \Psi = 4\pi G \rho \qquad \text{(Gravitationspotential)}$$

Bessere Modelle beinhalten innere Reibung (Navier-Stokes Gleichungen), erweiterte Zustandsgleichung und Strahlungstransport.

Existenz und Regularität der inkompressiblen Navier-Stokes-Gleichungen ist eines der sieben Millenium Prize Problems.

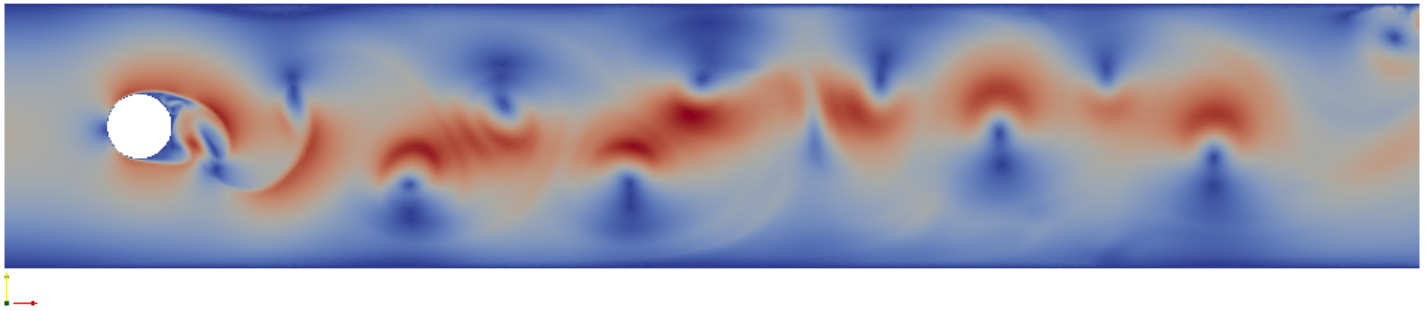
Numerische Simulation der Sternentstehung

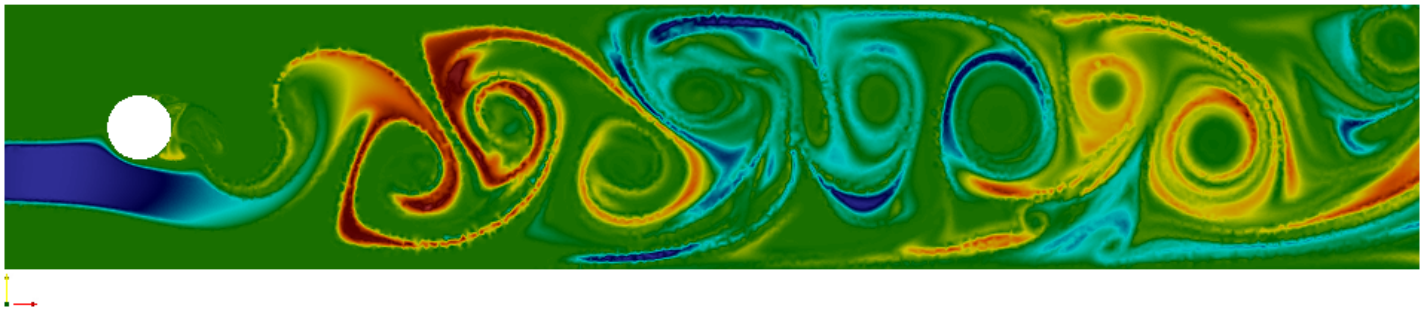
durchgeführt von Marvin Tegeler in seiner Diplomarbeit (2011).



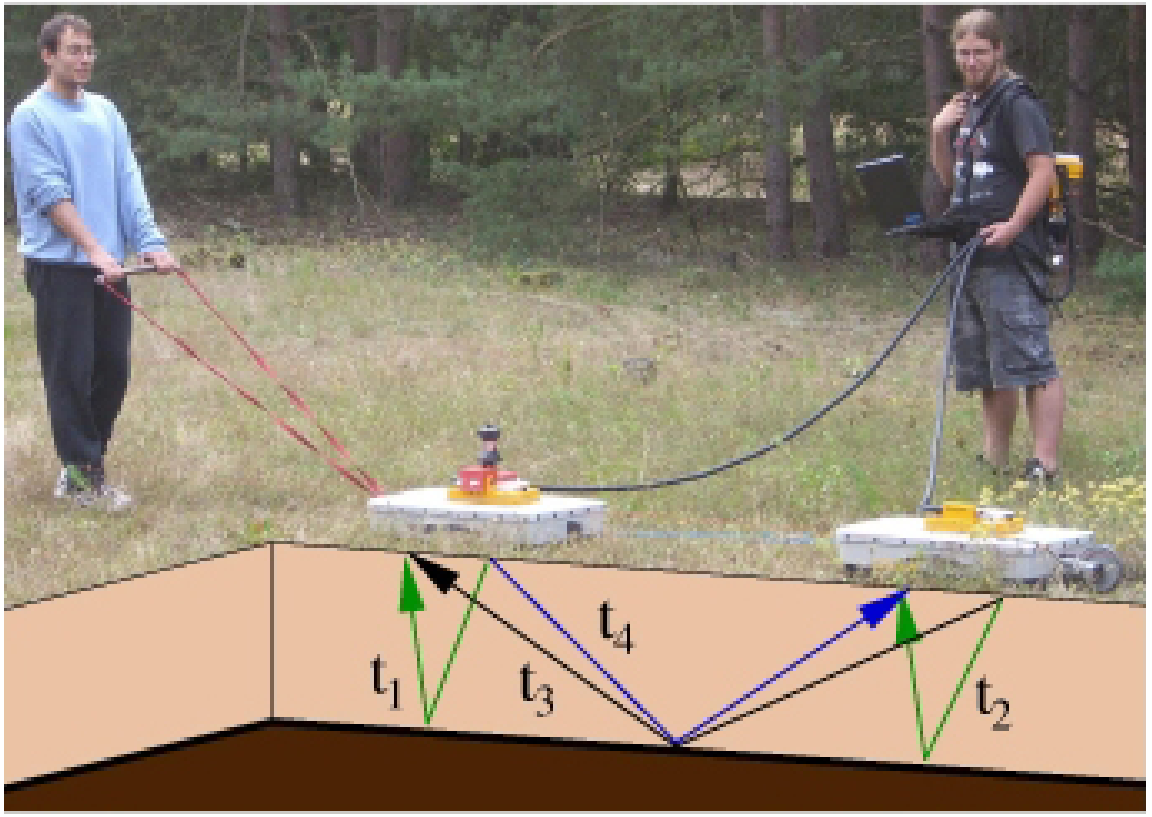
Von Karmannsche Wirbelstraße @ Re 1500

Diplomarbeit Marian Piatkowski





Bodenradar



Bestimme Strukturen im Boden durch Reflexion von Radarwellen

(Makroskopische) Maxwell-Gleichungen

beschreiben die Ausbreitung elektromagnetischer Wellen und wurden von James Clerk Maxwell im Jahr 1861 angegeben.

System linearer partieller Differentialgleichungen erster Ordnung:

$$\nabla \times E = -\partial_t B \qquad \text{(Faraday)}$$

$$\nabla \times H = j + \partial_t D \qquad \text{(Ampère)}$$

$$\nabla \cdot D = \rho \qquad \text{(Gauß)}$$

$$\nabla \cdot B = 0 \qquad \text{(Gauß für Magnetfeld)}$$

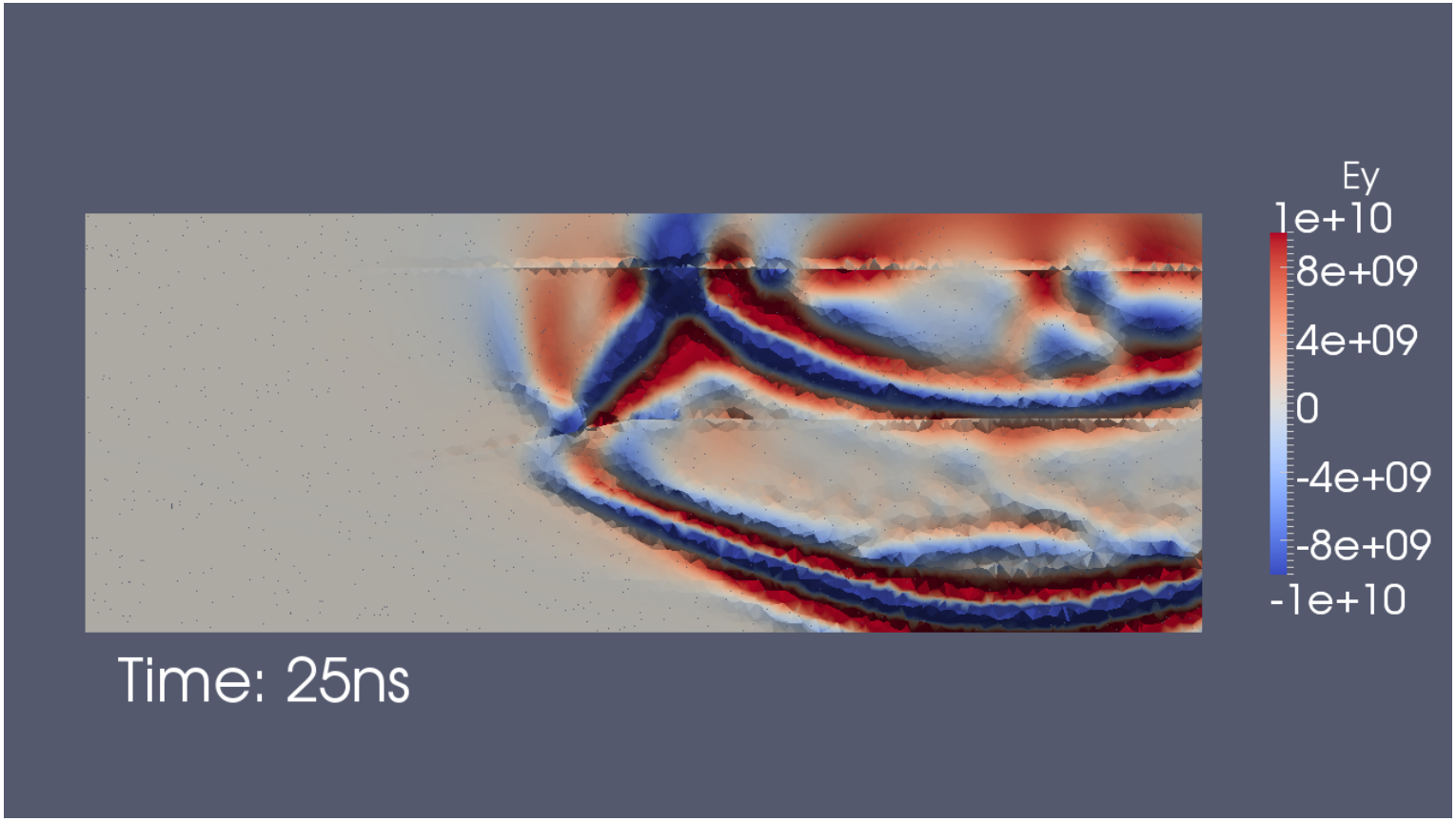
$$D = \epsilon_0 E + P \qquad \text{(elektrische Flussdichte)}$$

$$H = \mu_0^{-1} B - M \qquad \text{(magnetische Feldstärke)}$$

plus Rand- und Anfangsbedingungen

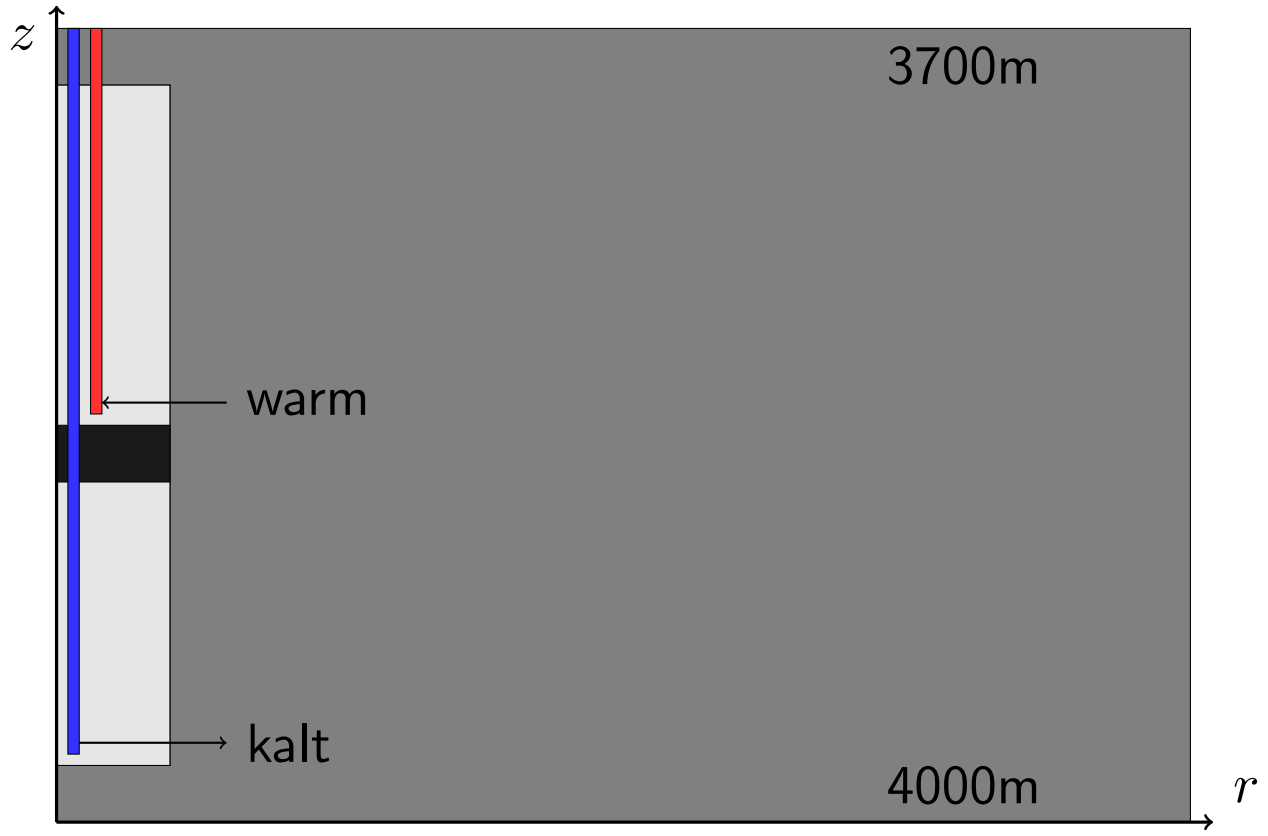
Simulation des Bodenradars

Jorrit Fahlke (IWR, 2011)



Eine Geothermieanlage

Einlochanlage in einem tiefen Aquifer (Zweidimensionaler Schnitt)



Gekoppelte Wasser- und Wärmeströmung

System nichtlinearer partieller DGL für Druck p und Temperatur T:

$$\partial_t(\phi\rho_w) + \nabla \cdot \{\rho_w u\} = f \qquad \text{(Massenerhaltung)}$$

$$u = \frac{k}{\mu}(\nabla p - \rho_w g) \qquad \text{(Darcy-Gesetz)}$$

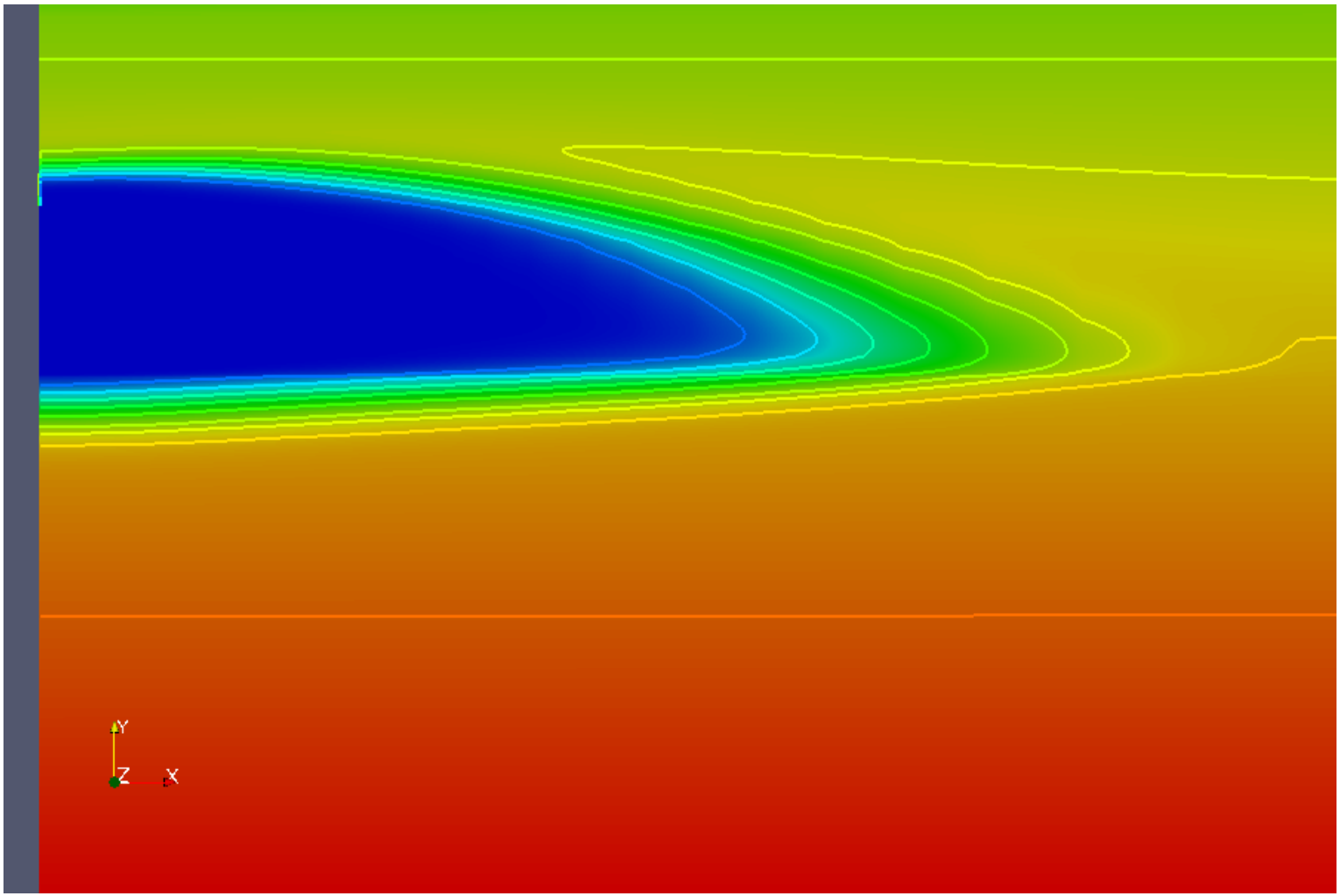
$$\partial_t(c_e\rho_e T) + \nabla \cdot q = g \qquad \text{(Energieerhaltung)}$$

$$q = c_w\rho_w u T - \lambda\nabla T \qquad \text{(Wärmefluss)}$$

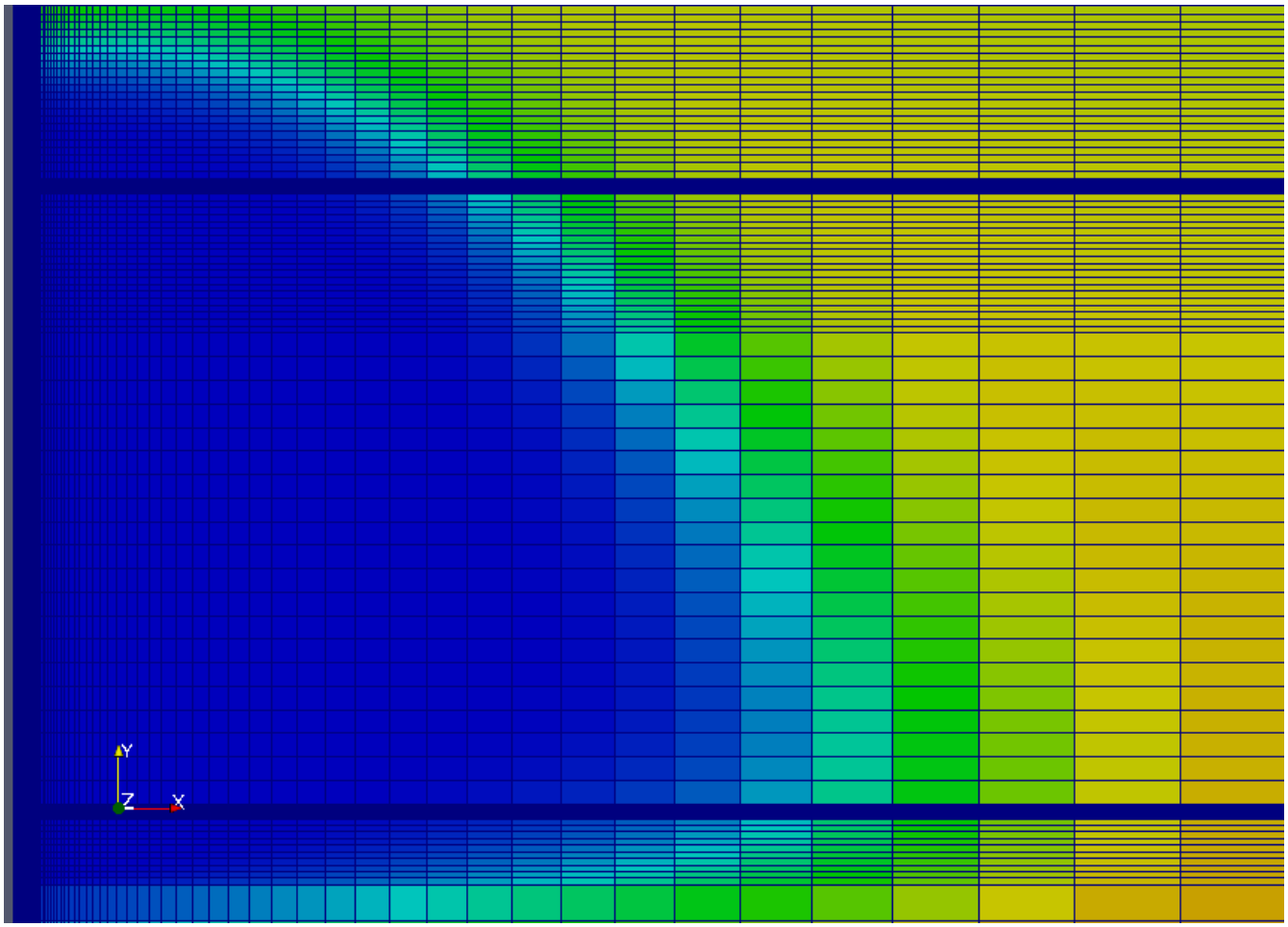
Nichtlinearität: $\rho_w(T)$, $\rho_e(T)$, $\mu(T)$

Permeabilität $k(x)$: 10^{-7} im kiesgefüllten Bohrloch, 10^{-16} im Verschluss

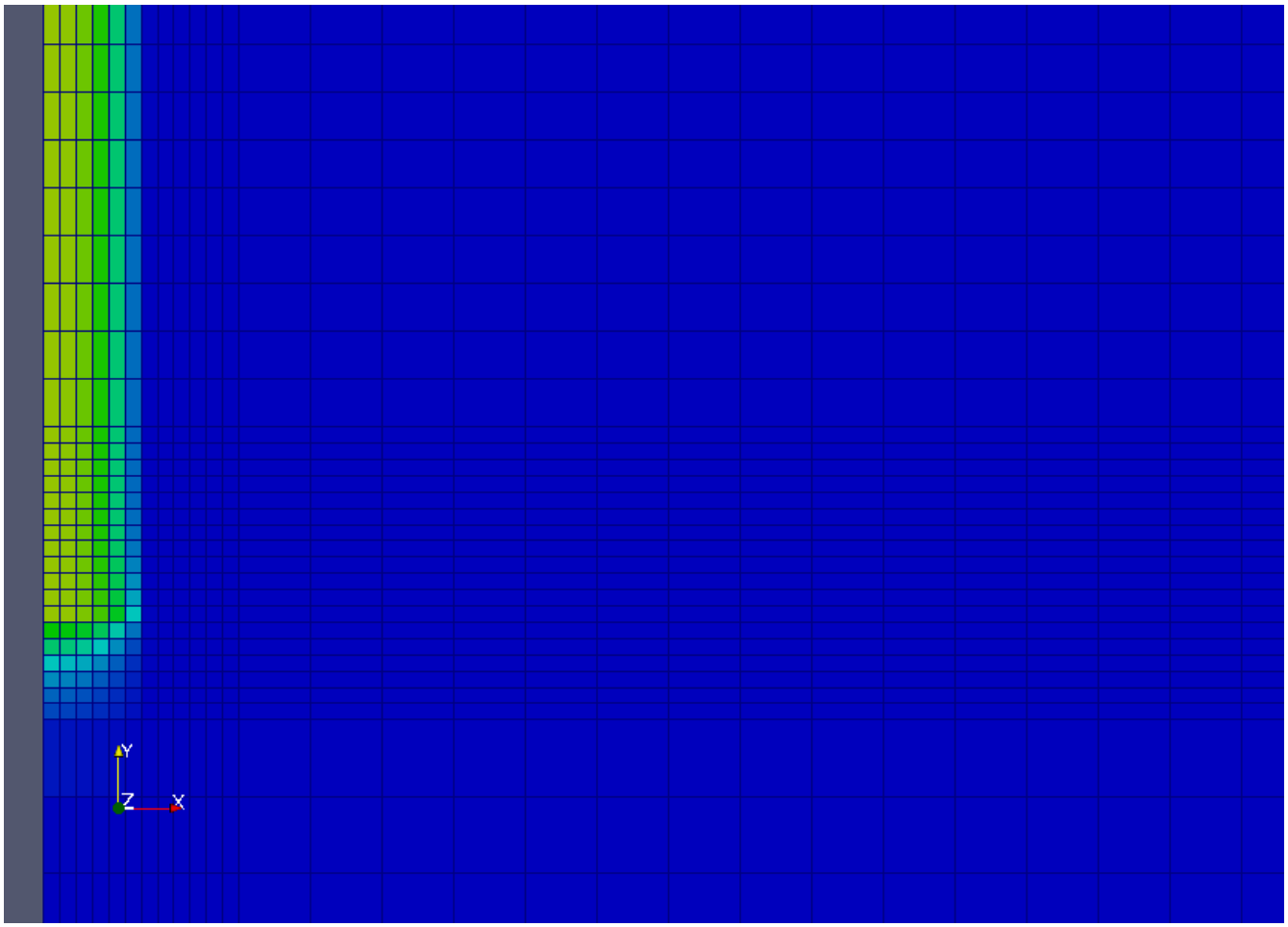
Raum-/Zeitskalen: R=15 km, r_b=14 cm, Sekunden (0.3 m/s im Bohrloch) bis Jahre



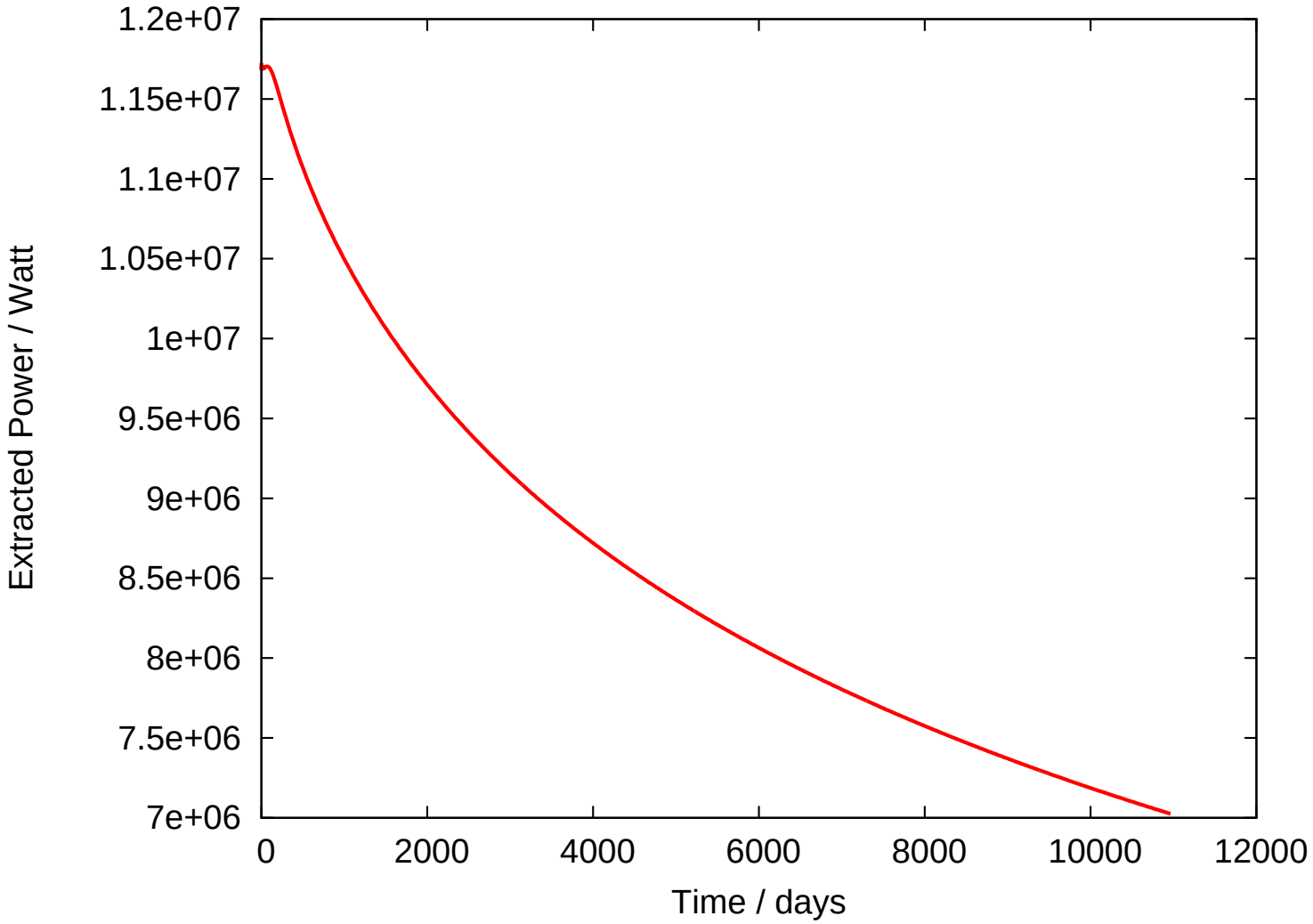
Temperaturverlauf nach 30 Jahren Betrieb



Detail am Einspeisebereich



Temperatur im Bohrloch



Entzugsleistung über 30 Jahre

Dichtegetriebene Strömung

in einem porösen Medium

$$\nabla \cdot v = 0, \qquad v = -(\nabla P - \omega_s \mathbf{1}_z)$$

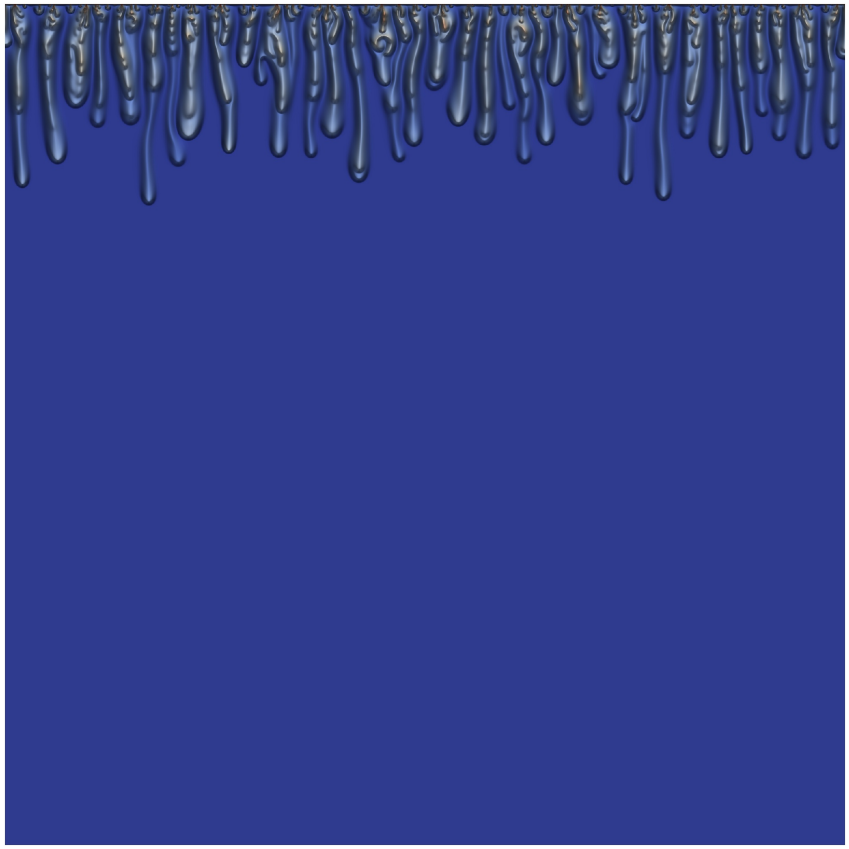
$$\partial_t \omega_s + \nabla \cdot \left(v\omega_s - \frac{1}{Ra} \nabla \omega_s \right) = 0$$

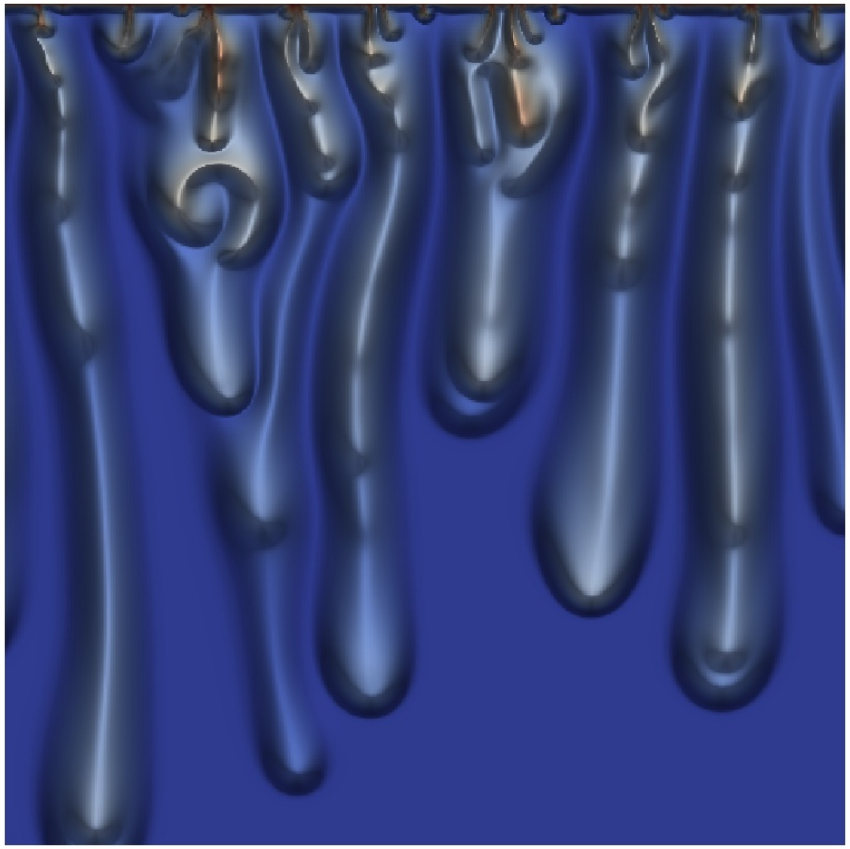
Dichteres Fluid über weniger dichtem Fluid führt zu instabiler Strömung

Erhöht die Durchmischung, z. B. bei der Sequestrierung von CO_2

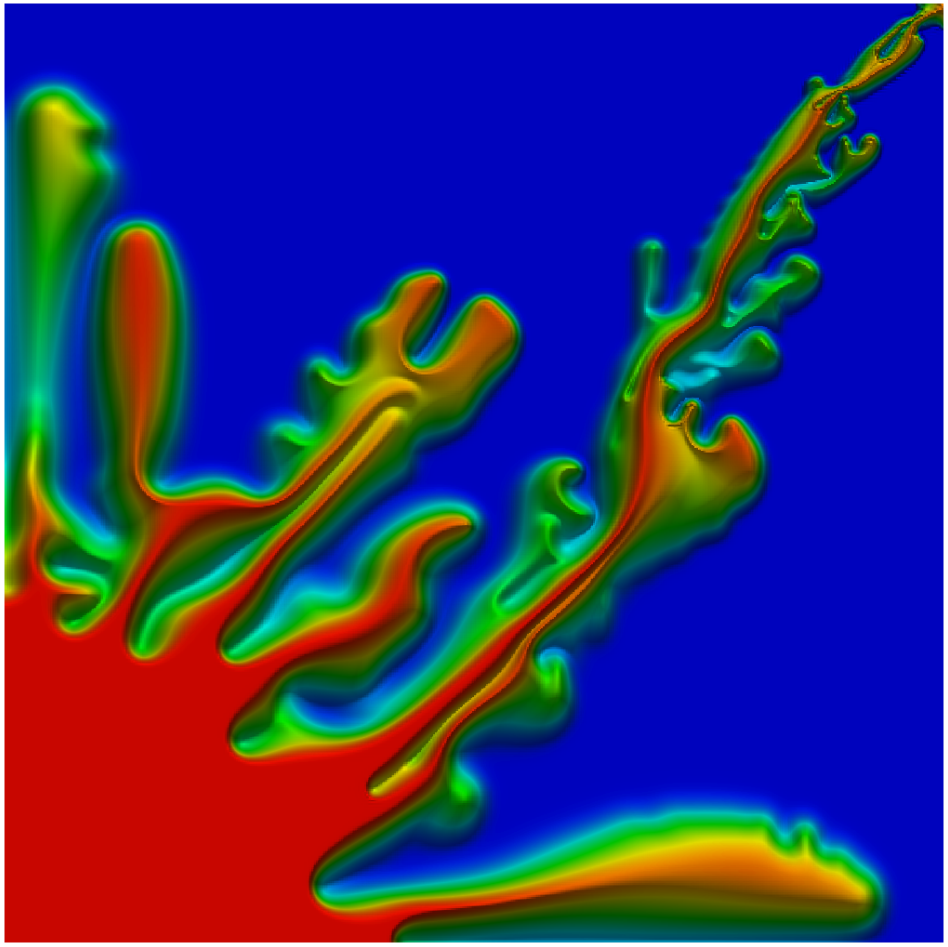
Wichtiger Effekt bei geophysikalischen Strömungen (dort: Navier-Stokes Gleichung)

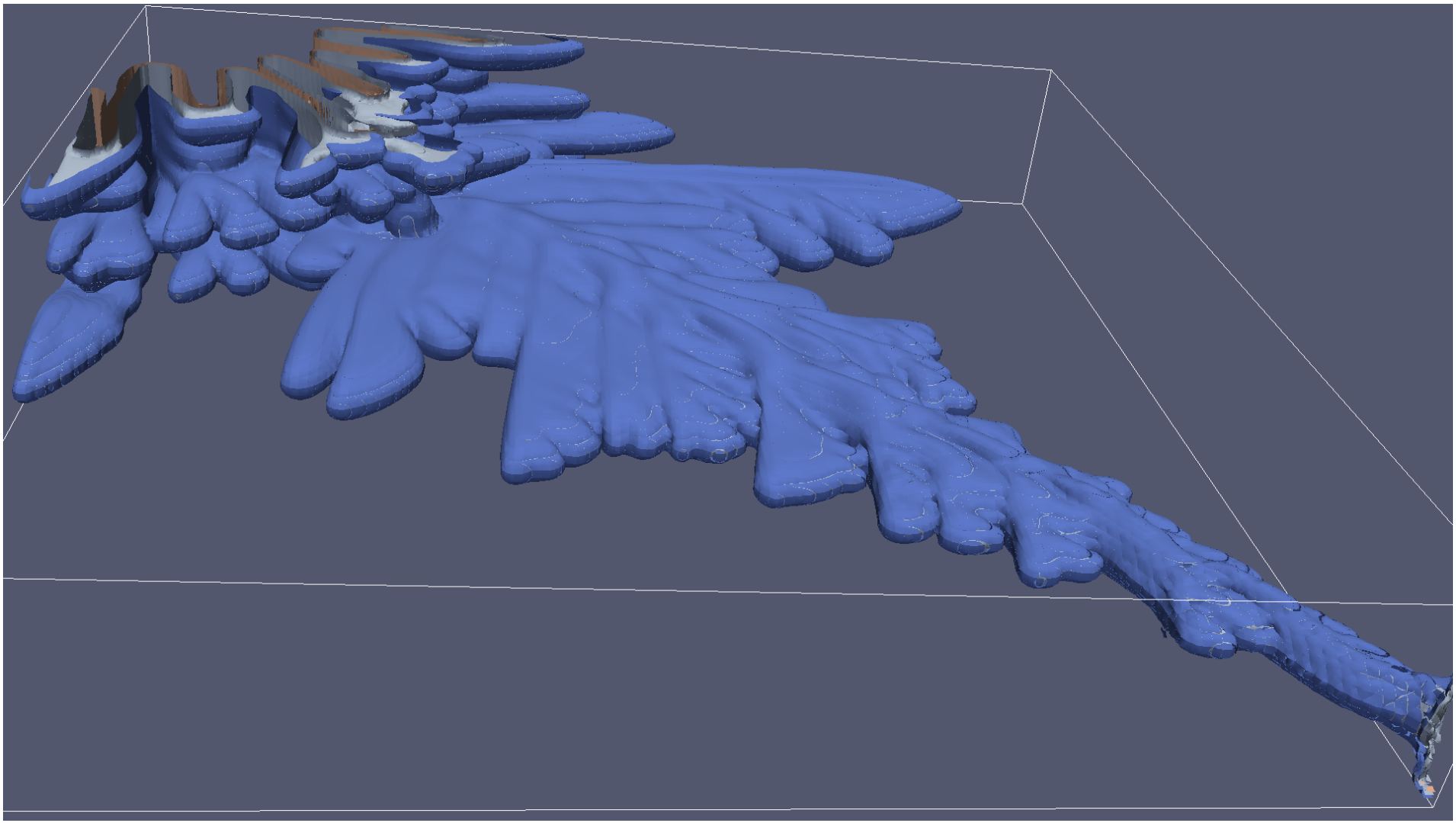
Zweidimensionale Simulation:





Viscous Fingering





Alle Simulationen wurden mit dem Softwarerahmenwerk DUNE durchgeführt

`http://www.dune-project.org`

Viele weitere Anwendungen

- Wetter und Klima
- Ölreservoirsimulation, CO_2 Sequestrierung, Lagerung radioaktiver Abfälle
- Geophysikalische Strömungen im Erdinnern
- Tsunamisimulation
- Festigkeit von Materialien
- Brennstoffzellen
- ...

Numerische Lösung partieller Differentialgleichungen

Physik des 19. Jahrhunderts, aber Mathematik und Informatik des 20. Jahrhunderts!

Treibende Kraft bei der Entwicklung des Computers, insbesondere Höchstleistungsrechner:

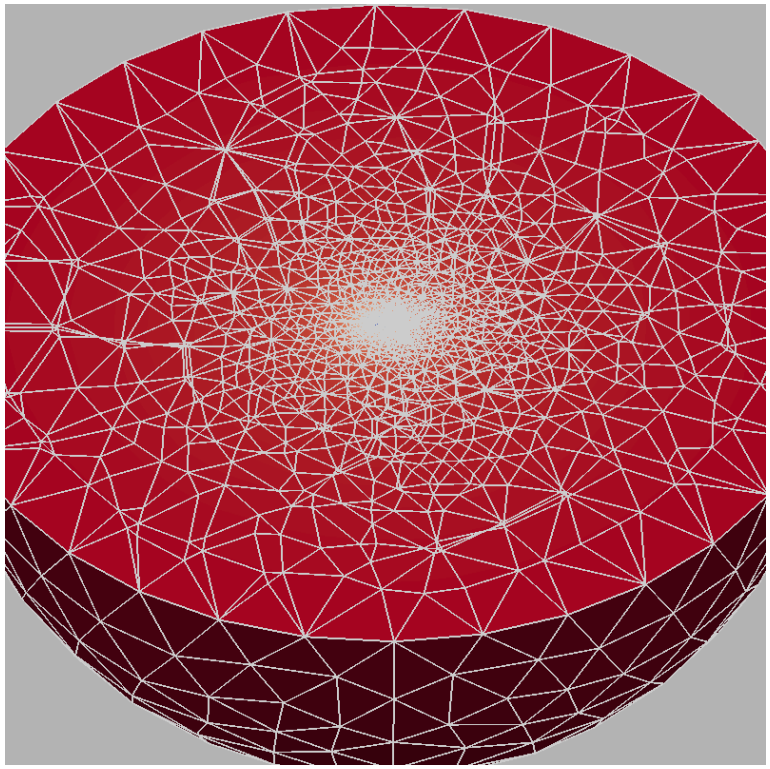
> An automatic computing system is a (usually highly composite) device, which can carry out instructions to perform calculations of a considerable order of complexity — e.g. to solve a non-linear partial differential equation in 2 or 3 independent variables numerically.
>
> John von Neumann, First Draft of a Report on the EDVAC, 30. Juni 1945

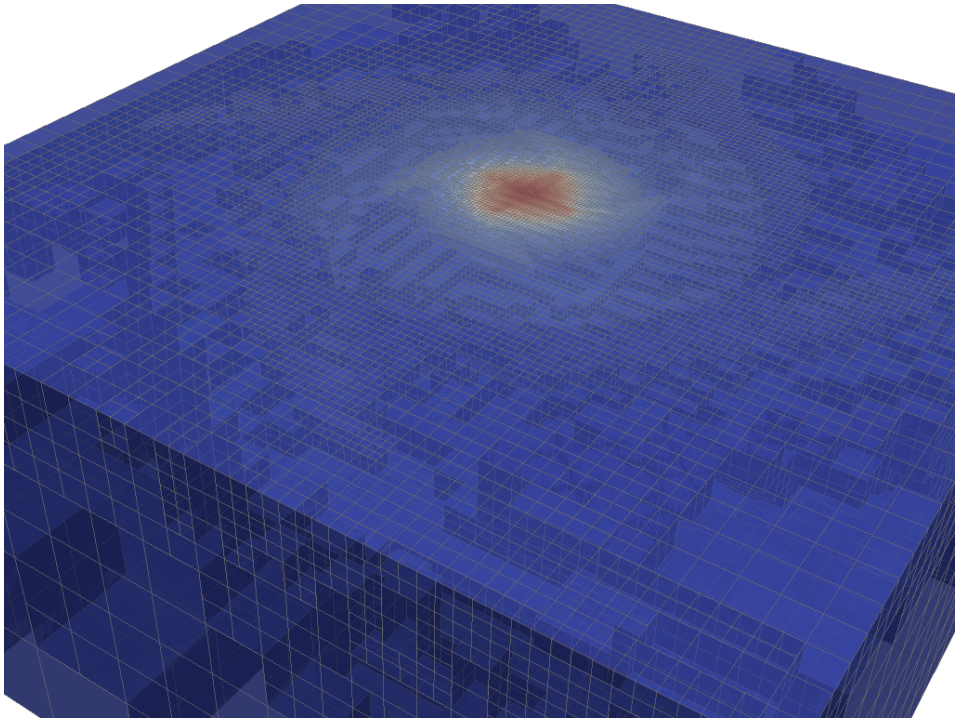
Verbindet Informatik, Physik und Mathematik

Etwa Jugene (294.912 Cores, Nummer 2 in Europa im Jahr 2011)

Ermöglicht z. B. die iterative Lösung von (bestimmten) linearen Gleichungssystemen mit 10^{11} Unbekannten in 4 Minuten.

Erfordert komplexe Algorithmen und Datenstrukturen, z. B. zur adaptiven Triangulierung:





Umfangreiches Softwareprojekt: `http://www.dune-project.org`

Literatur

H.-J. Appelrath, J. Ludewig: Skriptum Informatik – eine konventionelle Einführung. B. G. Teubner Verlag, 5. Auflage, 2000.

H. Abelson, G. J. Sussman mit J. Sussman: Struktur und Interpretation von Computerprogrammen, Springer Verlag, 1998.

B. Stroustrup: The C++ Programming Language, Addison-Wesley, 4. Auflage, 2013.

C++ Programming, freies WikiBook `http://en.wikibooks.org/wiki/C%2B%2B_Programming`

U. Schöning: Ideen der Informatik. Oldenburg Verlag, 2002.

D. R. Hofstatter: Gödel, Escher Bach: Ein Endloses Geflochtenes Band. dtv Taschenbuch, 11. Auflage, 2007.

Ankündigungen

- Hier nochmal der Link zur Vorlesung:
  `www.iwr.uni-heidelberg.de/groups/viscomp/teaching/2015-16/ipi/`

- Es wurde ein zusätzliches MÜSLI für die UNIX-Einführung und das Betreute Programmieren eingerichtet, bitte bis Fr 16.10. 16:00 Uhr eintragen.

- Bis Do 15.10. 23:59 Uhr können Sie noch Änderungen im Übungs-MÜSLI vornehmen.

- Übungsgruppe wechseln nach der Einteilung:
  Finden Sie einen Tauschpartner in der anderen Gruppe und senden Sie beide eine email an `ipi2015@iwr.uni-heidelberg.de`

- Heute erstes Übungsblatt

- Übungen beginnen am Mo 19.10.

- Abgabe des ersten Übungsblatts am Do 29.10., 14:15 Uhr (in zwei Wochen)

- Abgabe: Zettelkästen am INF 288, in der Ecke zwischen HS 2 und Seifertraum

Grundbegriffe

Diese Vorlesung: Ein kleiner Ausflug in die Theoretische Informatik.

Diese untersucht (neben anderen Dingen) die Frage ,,Was ist berechenbar"?

Dafür ist eine Formalisierung des Rechenbegriffes notwendig.

Inhalt:

- Textersetzungssysteme
- Graphen und Bäume
- Turingmaschine

Formale Systeme: MIU

Im folgenden betrachten wir Zeichenketten über einem Alphabet.

Ein Alphabet $\mathcal{A}$ ist eine endliche, nichtleere Menge (manchmal verlangt man zusätzlich, dass die Menge geordnet ist). Die Elemente von $\mathcal{A}$ nennen wir Zeichen (oder Symbole).

Eine endliche Folge nicht notwendigerweise verschiedener Zeichen aus $\mathcal{A}$ nennt man ein Wort. Das leere Wort ϵ besteht aus keinem einzigen Zeichen. Es ist ein Symbol für „Nichts“.

Die Menge aller möglichen Wörter inklusive dem leeren Wort wird als freies Monoid $\mathcal{A}^*$ bezeichnet.

Beispiel: $\{0,1\}^* = \{\epsilon, 0, 1, 00, 01, 10, 11, 000, \ldots\}$

Formale Systeme dienen der Beschreibung interessanter Teilmengen von $\mathcal{A}^*$.

Definition: Ein formales System ist ein System von Wörtern und Regeln. Die Regeln sind Vorschriften für die Umwandlung eines Wortes in ein anderes.

Mathematisch: $F = (\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{R})$, wobei

- $\mathcal{A}$ das Alphabet,
- $\mathcal{B} \subseteq \mathcal{A}^*$ die Menge der wohlgebildeten Worte,
- $\mathcal{X} \subset \mathcal{B}$ die Menge der Axiome und
- $\mathcal{R}$ die Menge der Produktionsregeln

sind. Ausgehend von $\mathcal{X}$ werden durch Anwendung von Regeln aus $\mathcal{R}$ alle wohlgeformten Wörter $\mathcal{B}$ erzeugt.

Formale Systeme entstanden Anfang des 20. Jahrhunderts im Rahmen der Formalisierung der Mathematik. Ziel war es ein System zu schaffen mit dem alle

mathematischen Sätze (wahre Aussagen über einen mathematischen Sachverhalt, möglicherweise in Teilgebieten der Mathematik) aus einem kleinen Satz von Axiomen mittels Regeln hergeleitet werden können (Hilbertprogramm[1]).

Wir betrachten hier formale System nur im Sinne „formaler Sprachen“, die später noch ausführlicher behandelt werden.

[1] David Hilbert, dt. Mathematiker, 1862–1943.

Beispiel: MIU-System (aus [Hofstadter[2], 2007])

Das MIU-System handelt von Wörtern, die nur aus den drei Buchstaben M, I, und U bestehen.

- $\mathcal{A}_{\mathrm{MIU}} = \{\mathrm{M}, \mathrm{I}, \mathrm{U}\}$.
- $\mathcal{X}_{\mathrm{MIU}} = \{\mathrm{MI}\}$.
- $\mathcal{R}_{\mathrm{MIU}}$ enthält die Regeln:

1. $\mathrm{M}x\mathrm{I} \to \mathrm{M}x\mathrm{IU}$. Hierbei ist $x \in \mathcal{A}^*_{\mathrm{MIU}}$ irgendein Wort oder ϵ.
   Beispiel: MI $\to$ MIU. Man sagt MIU wird aus MI abgeleitet.
2. $\mathrm{M}x \to \mathrm{M}xx$.
   Beispiele: MI $\to$ MII, MIUUI $\to$ MIUUIIUUI.
3. $x\mathrm{III}y \to x\mathrm{U}y$ $(x, y \in \mathcal{A}^*_{\mathrm{MIU}})$.
   Beispiele: MIII $\to$ MU, UIIIIM $\to$ UUIM, UIIIIM $\to$ UIUM.
4. $x\mathrm{UU}y \to xy$.
   Beispiele: UUU $\to$ U, MUUUIII $\to$ MUIII.

[2] Douglas R. Hofstadter, US-amerk. Physiker, Informatiker und Kognitionswissenschaftler, geb. 1945.

- $\mathcal{B}_{\mathrm{MIU}}$ sind dann alle Worte die ausgehend von den Elementen von $\mathcal{X}$ mithilfe der Regeln aus $\mathcal{R}$ erzeugt werden können, also

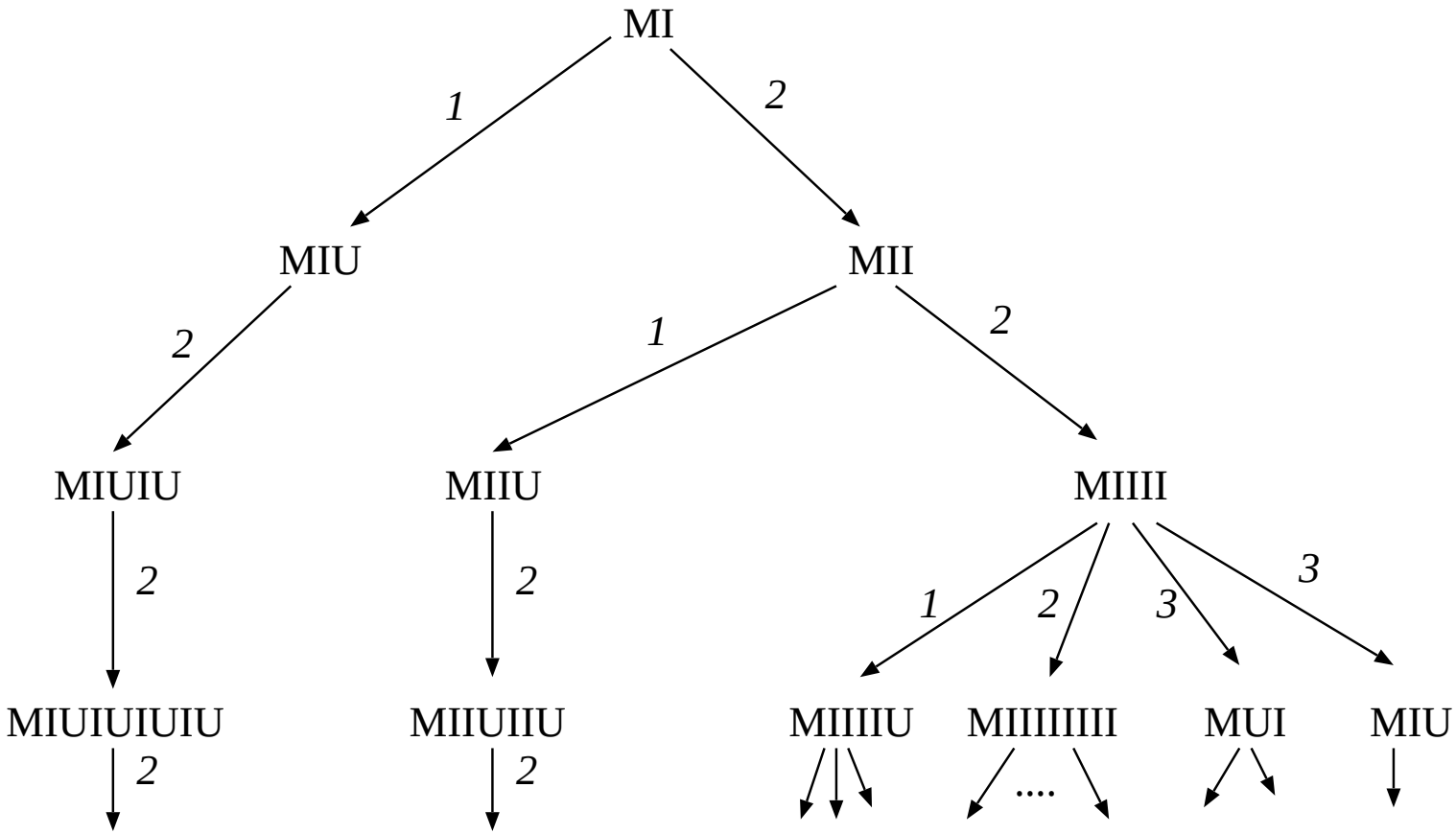
$$\mathcal{B} = \{\mathrm{MI}, \mathrm{MIU}, \mathrm{MIUUI}, \ldots\}.$$

Beobachtung: B_{MIU} enthält offenbar unendlich viele Worte.

Problem: (MU-Rätsel) Ist MU ein Wort des MIU-Systems?
Oder mathematisch: $\mathrm{MU} \in \mathcal{B}_{\mathrm{MIU}}$?

Systematische Erzeugung aller Worte des MIU-Systems

Dies führt auf folgende Baumstruktur:



Beschreibung: Ganz oben steht das Anfangswort MI. Auf MI sind nur die Regeln 1 und 2 anwendbar. Die damit erzeugten Wörter stehen in der zweiten Zeile. Ein Pfeil bedeutet, dass ein Wort aus dem anderen ableitbar ist. Die Zahl an dem Pfeil ist die Nummer der angewendeten Regel. In der dritten Zeile stehen alle Wörter, die durch Anwendung von zwei Regeln erzeugt werden können, usw.

Bemerkung: Wenn man den Baum in dieser Reihenfolge durchgeht (Breitendurchlauf), so erzeugt man nach und nach alle Wörter des MIU-Systems.

Folgerung: Falls $\mathrm{MU} \in \mathcal{B}_{\mathrm{MIU}}$, wird dieses Verfahren in endlicher Zeit die Antwort liefern. Wenn dagegen $\mathrm{MU} \notin \mathcal{B}_{\mathrm{MIU}}$, so werden wir es mit obigem Verfahren nie erfahren!

Sprechweise: Man sagt: Die Menge $\mathcal{B}_{\mathrm{MIU}}$ ist rekursiv aufzählbar.

Frage: Wie löst man nun das MU-Rätsel?

Lösung des MU-Rätsels

Zur Lösung muss man Eigenschaften der Wörter in $\mathcal{B}_{\mathrm{MIU}}$ analysieren.

Beobachtung: Alle Ketten haben immer M vorne. Auch gibt es nur dieses eine M, das man genausogut hätte weglassen können. Hofstadter wollte aber das Wort MU herausbekommen, das in Zen-Koans eine Rolle spielt:

> Ein Mönch fragte einst Meister Chao-chou:
> „Hat ein Hund wirklich Buddha-Wesen oder nicht?“
> Chao-chou sagte: „Mu.“

Beobachtung: Die Zahl der I in einzelnen Worten ist niemals ein Vielfaches von 3, also auch nicht 0.

Beweis: Ersieht man leicht aus den Regeln, sei $\mathrm{anzahli}(n)$ die Anzahl der I nach Anwendung von n Regeln, $n \in \mathbb{N}_0$. Dann gilt:

$$\mathrm{anzahli}(n) = \begin{cases} 1 & n = 0, \text{Axiom}, \\ \mathrm{anzahli}(n-1) & n > 0, \text{Regel 1, 4}, \\ \mathrm{anzahli}(n-1) \cdot 2 & n > 0, \text{Regel 2}, \\ \mathrm{anzahli}(n-1) - 3 & n > 0, \text{Regel 3} \end{cases}$$

Ist $\mathrm{anzahli}(n-1) \bmod 3 \neq 0$, so gilt dies auch nach Anwendung einer beliebigen Regel.

Von Graphen und Bäumen

Der Baum ist eine sehr wichtige Struktur in der Informatik und ein Spezialfall eines Graphen.

Definition: Ein Graph $G = (V, E)$ besteht aus

- einer nichtleeren Menge V, der sogenannten Menge der Knoten, sowie
- der Menge der Kanten $E \subseteq V \times V$.

$V \times V = \{(v, w) \ : \ v, w \in V\}$ bezeichnet das kartesische Produkt.

Teilmengen von $V \times V$ bezeichnet man auch als Relationen.

Beispiel: Gleichheit als Relation. Sei V eine Menge (dies impliziert, dass alle Elemente verschieden sind). Setze

$$E_= = \{(v, w) \in V \times V \ : \ v = w\}.$$

Dann gilt $v = w \Leftrightarrow (v, w) \in E_{=}$.

Wichtige Spezialfälle von Graphen sind:

- Ungerichter Graph: $(v, w) \in E \Rightarrow (w, v) \in E$. Sonst heisst der Graph gerichtet.
- Verbundener Graph: Ein ungerichteter Graph heisst verbunden, falls jeder Knoten von jedem anderen Knoten über eine Folge von Kanten erreichbar ist. Bei einem gerichteten Graphen ergänze erst alle Kanten der Gegenrichtung und wende dann die Definition an.
- Zyklischer Graph: Es gibt, ausgehend von einem Knoten, eine Folge von Kanten mit der man wieder beim Ausgangsknoten landet.

Definition: Wir definieren die Menge der Bäume rekursiv über die Anzahl der Knoten als Teilmenge aller möglicher Graphen.

- $(\{v\}, \emptyset)$ ist ein Baum.

- Sei $B = (V, E)$ ein Baum, so ist $B' = (V', E')$ ebenfalls ein Baum, wenn

$$V' = V \cup \{v\}, \qquad v \notin V, \qquad E' = E \cup \{(w, v) : w \in V\}.$$

  Man hängt also einen neuen Knoten an genau einen Knoten des existierenden Baumes an. v heisst Kind und w wollen wir geschlechtsneutral als Elter von v bezeichnen.

Bemerkung: Auch andere Definitionen sind möglich, etwa als zyklenfreier, verbundener Graph.

Bezeichnung:

- Jeder Baum besitzt genau einen Knoten, der keine eingehenden Kanten hat. Dieser heisst Wurzel.

- Knoten ohne ausgehende Kanten heissen Blätter, alle anderen Knoten heissen innere Knoten

- Ein Baum bei dem jeder innere Knoten höchstens zwei Kinder hat heisst Binärbaum.

Beobachtung: Ein Baum ist verbunden. Es gibt genau einen Weg von der Wurzel zu jedem Blatt.

Turingmaschine

Als weiteres Beispiel für ein „Regelsystem“ betrachten wir die Turingmaschine (TM).

Diese wurde 1936 von Alan Turing[3] zum theoretischen Studium der Berechenbarkeit eingeführt.

Wissen: Der sogenannte Turing-Preis (Turing Award) ist so etwas wie der „Nobelpreis der Informatik“.

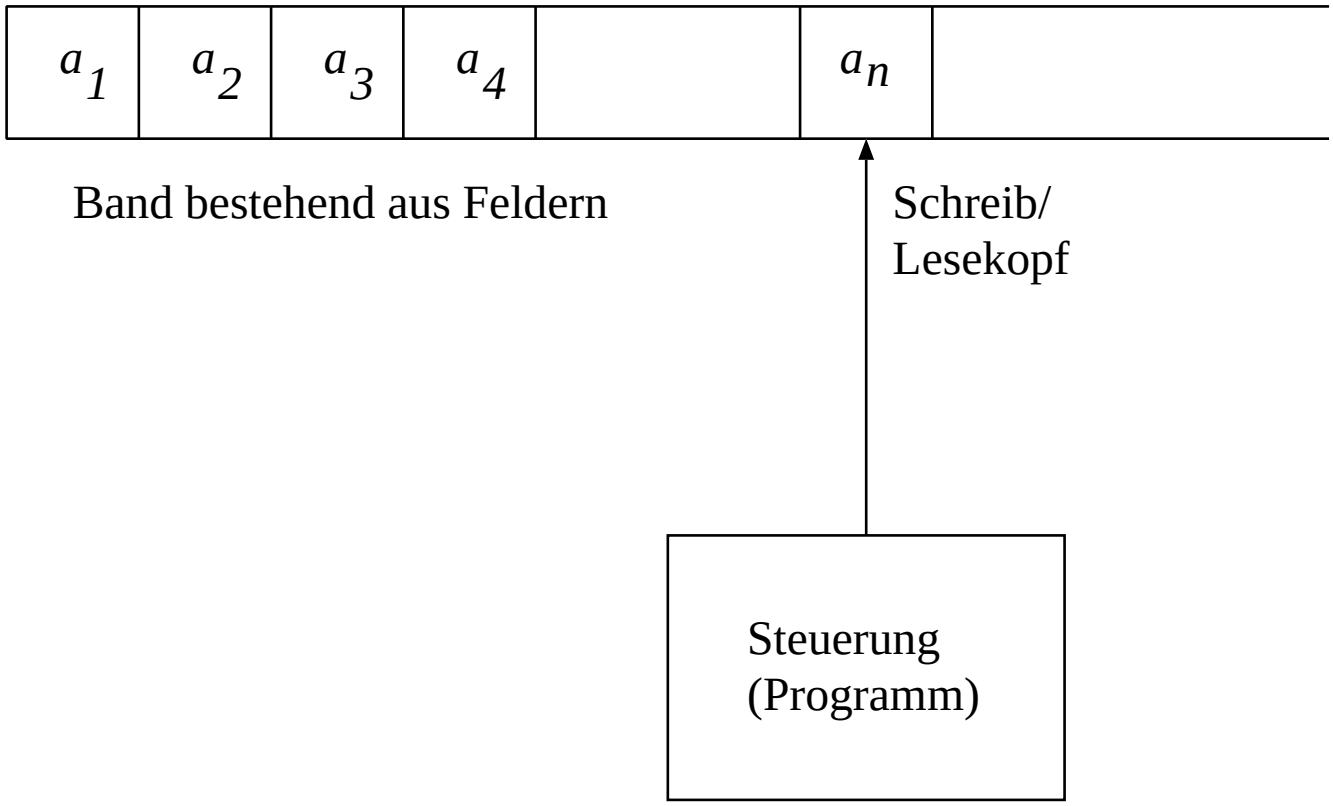
Eine TM besteht aus einem festen Teil („Hardware“) und einem variablen Teil („Software“). TM bezeichnet somit nicht eine Maschine, die genau eine Sache tut, sondern ist ein allgemeines Konzept, welches eine ganze Menge von verschiedenen Maschinen definiert. Alle Maschinen sind aber nach einem festen Schema aufgebaut.

Die **Hardware** besteht aus einem einseitig unendlich langen Band welches aus einzelnen Feldern besteht, einem Schreib-/Lesekopf und der Steuerung. Jedes Feld

[3] Alan Turing, brit. Mathematiker, 1912–1954.

des Bandes trägt ein Zeichen aus einem frei wählbaren (aber für eine Maschine festen) Bandalphabet (Menge von Zeichen). Der Schreib-/Lesekopf ist auf ein Feld positioniert, welches dann gelesen oder geschrieben werden kann. Die Steuerung enthält den variablen Teil der Maschine und wird nun beschrieben.

Diese Beschreibung suggeriert, dass eine TM als eine Art primitiver Computer verstanden werden kann. Dies war aber nicht die Absicht von Alan Turing. Er verstand diese als Gedankenmodell um die Berechenbarkeit von Funktionen zu studieren.


a_1
a_2
a_3
a_4
a_n
Band bestehend aus Feldern
Schreib/
Lesekopf
Steuerung
(Programm)

Die Steuerung, der variable Teil der Maschine, befindet sich in einem von endlich vielen Zuständen und arbeitet wie folgt:

1. Am Anfang befindet sich die Maschine im sog. Startzustand, das Band ist mit einer Eingabe belegt und die Position des Schreib-/Lesekopfes ist festgelegt.

2. Lese das Zeichen unter dem Lesekopf vom Band.

3. Abhängig vom gelesenen Zeichen und dem aktuellen Zustand der Steuerung führe **alle** folgende Aktionen aus:

   - Schreibe ein Zeichen auf das Band,
   - bewege den Schreib-/Lesekopf um ein Feld nach links oder rechts,
   - überführe die Steuerung in einen neuen Zustand.

4. Wiederhole diese Schritte solange bis ein spezieller Endzustand erreicht wird.

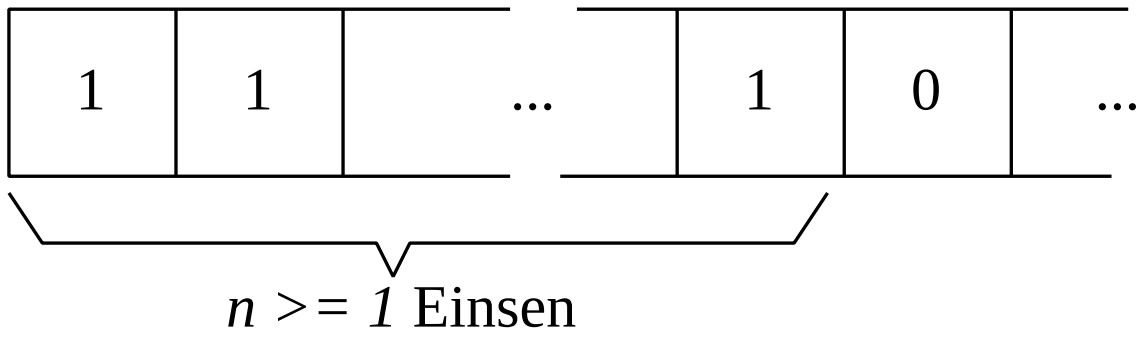
Die auszuführenden Aktionen kann man in einer Übergangstabelle notieren. Diese Tabelle nennt man auch Programm.

Beispiel:

Zustand	Eingabe	Operation	Folgezustand
1	0	0,links	2
2	1	1,rechts	1

Jede Zeile der Tabelle beschreibt die auszuführenden Aktionen für eine Eingabe/Zustand-Kombination. Links vom Doppelbalken stehen Eingabe und Zustand, rechts davon Ausgabe, Bewegungsrichtung und Folgezustand.

Beispiel: Löschen einer Einserkette. Das Bandalphabet enthalte nur die Zeichen 0 und 1. Zu Beginn der Bearbeitung habe das Band folgende Gestalt:



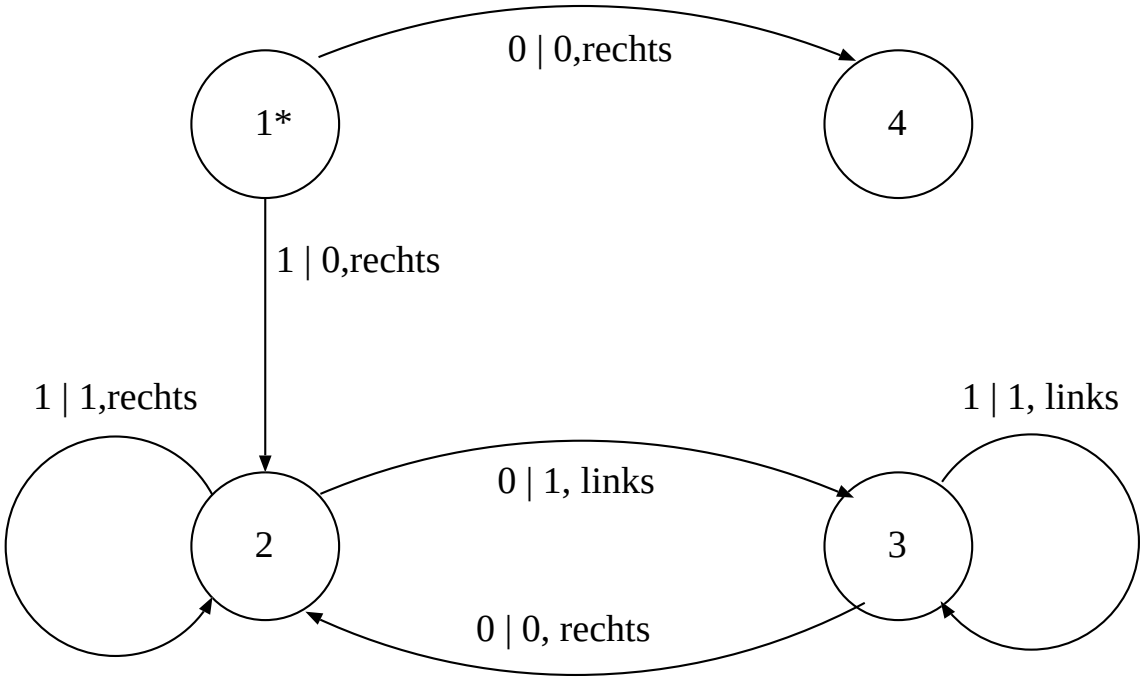
Der Kopf steht zu Beginn auf der Eins ganz links. Folgendes Programm mit zwei Zuständen löscht die Einserkette und stoppt:

Zustand	Eingabe	Operation	Folgezustand	Bemerkung
1	1	0,rechts	1	Anfangszustand
	0	0,rechts	2	
2				Endzustand

Beispiel: Raten Sie was folgendes Programm macht:

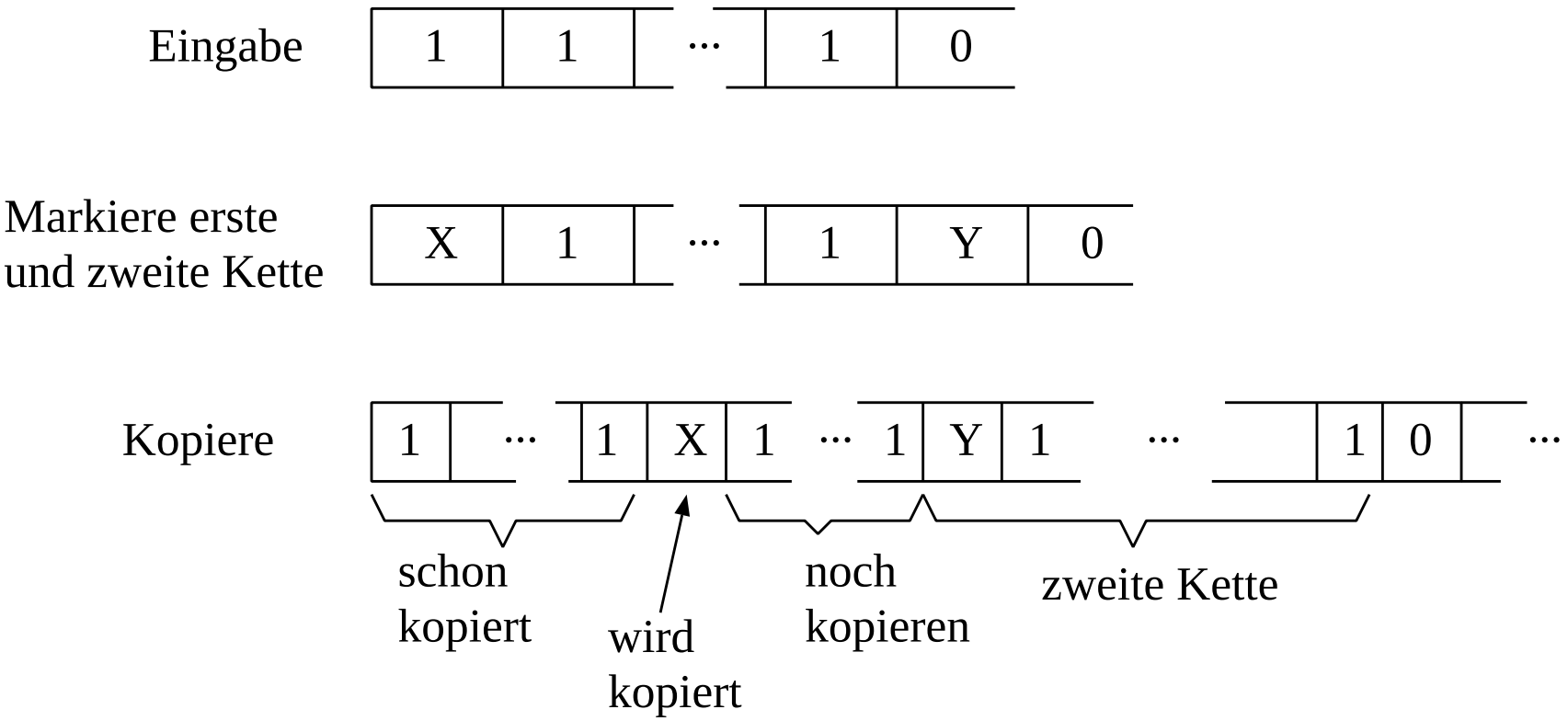
Zustand	Eingabe	Operation	Folgezustand	Bemerkung
1	1	0,rechts	2	Anfangszustand
	0	0,rechts	4	
2	1	1,rechts	2	
	0	1,links	3	
3	1	1,links	3	
	0	0,rechts	2	
4				Endzustand

TM-Programme lassen sich übersichtlicher als Übergangsgraph darstellen. Jeder Knoten ist ein Zustand. Jeder Pfeil entspricht einer Zeile der Tabelle. Hier das Programm des vorigen Beispiels als Graph:

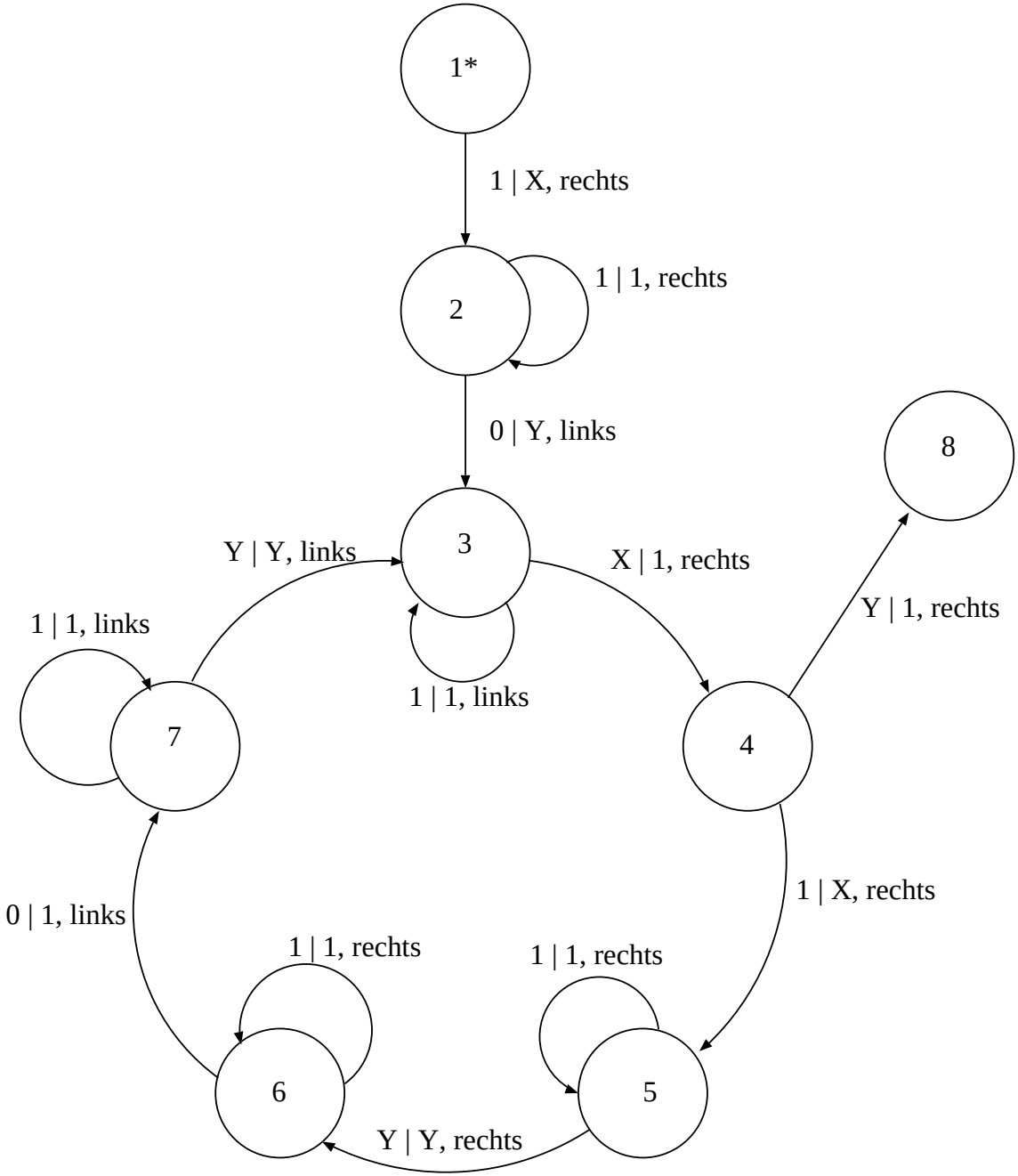


Beispiel: Verdoppeln einer Einserkette. Eingabe: n Einsen wie in Beispiel 1. Am Ende der Berechnung sollen ganz links $2n$ Einsen stehen, sonst nur Nullen.

Wie löst man das mit einer TM? Hier eine Idee:



Das komplette Programm ist schon ganz schön kompliziert und sieht so aus:


1*
1 | X, rechts
2
1 | 1, rechts
0 | Y, links
8
3
Y | Y, links
X | 1, rechts
Y | 1, rechts
1 | 1, links
1 | 1, links
7
4
1 | X, rechts
0 | 1, links
1 | 1, rechts
1 | 1, rechts
6
5
Y | Y, rechts

Bemerkung: Wir erkennen die drei wesentlichen Komponenten von Berechnungsprozessen:

- Grundoperationen
- Selektion
- Wiederholung

Ankündigungen

- Übungs-Einteilung: Zufriedenheitsindex >99% !
- Einteilung UNIX-Einführung / Betreutes Programmieren: momentan fünf Termine, Zuteilung nicht einfach, aber scheint zu konvergieren. Bitte E-Mail an ipi2015@iwr.uni-heidelberg.de falls Sie noch keinen Platz zugewiesen bekommen haben.
- Skripte werden in ca. zwei Wochen bei der Fachschaft abzuholen sein (bzw. in der Vorlesung ausgeteilt).
- Anerkennung von Klausurzulassungen: Seit einem Jahr sind erworbene Zulassungen nur noch für ein Jahr gültig! Voraussetzung ist weiterhin, dass überhaupt ein Prüfungsversuch durchgeführt wurde.
- Herr Bastian hat uns seine Zulassungen mitgeteilt.

Stoff für heute

- Was ist ein Algorithmus
- Berechenbarkeit
- Reale Computer
- Programmiersprachen

Problem, Algorithmus, Programm

Definition: Ein Problem ist eine zu lösende Aufgabe. Wir sind daran interessiert Verfahren zu finden, die Aufgaben in einer Klasse von Problemen zu lösen. Das konkrete zu lösende Problem wird mittels Eingabeparameter ausgewählt.

Beispiel: Finde die kleinste von $n \geq 1$ Zahlen $x_1, \ldots, x_n$, $x_i \in \mathbb{N}$.

Definition: Ein Algorithmus beschreibt, wie ein Problem einer Problemklasse mittels einer Abfolge bekannter Einzelschritte gelöst werden kann. Beispiele aus dem Alltag, wie Kochrezepte oder Aufbauanleitungen für Abholmöbel erinnern an Algorithmen, sind aber oft nicht allgemein und unpräzise.

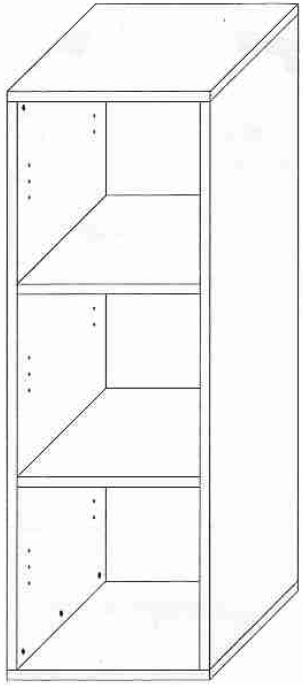
Beispiel: Das Minimum von n Zahlen könnte man so finden: Setze $\min = x_1$. Falls $n = 1$ ist man fertig. Ansonsten teste der Reihe nach für $i = 2, 3, \ldots, n$ ob $x_i < \min$. Falls ja, setze $\min = x_i$.

Ein Algorithmus muss gewisse Eigenschaften erfüllen:

- Ein Algorithmus beschreibt ein generelles Verfahren zur Lösung einer Schar von Problemen.
- Trotzdem soll die Beschreibung des Algorithmus endlich sein. Nicht erlaubt ist also z. B. eine unendlich lange Liste von Fallunterscheidungen.
- Ein Algorithmus besteht aus einzelnen Elementaroperationen, deren Ausführung bekannt und endlich ist. Als Elementaroperationen sind also keine „Orakel“ erlaubt.

AUFBAUANLEITUNG

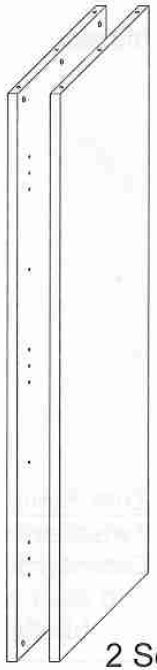
WÜRFELSYSTEM XERO



Aufbauanleitung am Beispiel eines Würfelsystems XERO in 116 cm Höhe mit 3 Zwischenböden.

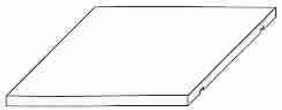
Stückliste

Die Anzahl der Zwischenböden, sowie die Maße der Seitenteile richtet sich nach Ihrer Bestellung.

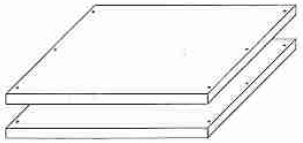


2 Seitenteile

Mögliche Höhen: 36 / 74 / 112 / 150 oder 188 cm



Zwischenboden



Satz Boden/Deckel

Zubehör

pro Satz Boden/Deckel

12x Verbinder

12x Verbinderbolzen

12x Gewindestift

1x Imbusschlüssel

pro Zwischenboden

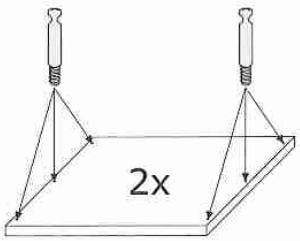
4x Excenterbolzen

4x Excenter

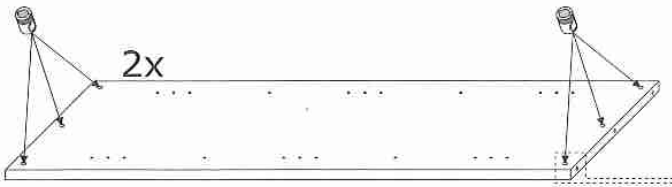
Tipp: Fetten oder ölen Sie die Gewinde der Bolzen vorher leicht ein, um das Eindrehen in das Massivholz zu erleichtern!

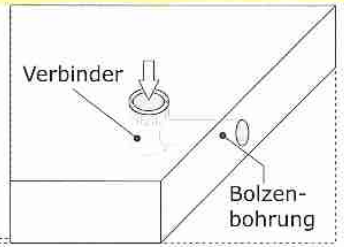
Bitten beachten Sie, dass ab einer Seitenhöhe von 150 cm zur zusätzlichen Stabilisierung mindestens ein Zwischenboden montiert werden muss!

1. Je sechs Verbinderbolzen mit einem geeigneten Schraubendreher in die Bohrungen von Boden und Deckel einschrauben.



2. Die Verbinder in die seitlichen Bohrungen der Seitenteile einstecken, so dass die seitliche Bohrung des Verbinders in Richtung der Bolzenbohrung weist.





Bemerkung: Spezielle Algorithmen sind:

- Terminierende Algorithmen: Der Algorithmus stoppt für jede zulässige Eingabe nach endlicher Zeit.

- Deterministische Algorithmen: In jedem Schritt ist bekannt, welcher Schritt als nächstes ausgeführt wird.

- Determinierte Algorithmen: Algorithmus liefert bei gleicher Eingabe stets das gleiche Ergebnis. Ein terminierender, deterministischer Algorithmus ist immer determiniert. Terminierende, nichtdeterministische Algorithmen können determiniert sein oder nicht.

Definition: Ein Programm ist eine Formalisierung eines Algorithmus. Ein Programm kann auf einer Maschine (z. B. TM) ausgeführt werden.

Beispiel: Das Minimum von n Zahlen kann mit einer TM berechnet werden. Die

Zahlen werden dazu in geeigneter Form kodiert (z. B. als Einserketten) auf das Eingabeband geschrieben.

Wir haben also das Schema: Problem $\Longrightarrow$ Algorithmus $\Longrightarrow$ Programm.

Die Informatik beschäftigt sich damit algorithmische Problemlösungen systematisch zu finden:

- Zunächst muss das Problem analysiert und möglichst präzise formuliert werden. Dieser Schritt wird auch als Modellierung bezeichnet.
- Im folgenden entwirft man einen effizienten Algorithmus zur Lösung des Problems. Dieser Schritt ist von zentralem Interesse für die Informatik.
- Schließlich muss der Algorithmus als Computerprogramm formuliert werden, welches auf einer konkreten Maschine ausgeführt werden kann.

Berechenbarkeit und Turing-Äquivalenz

Es sei $\mathcal{A}$ das Bandalphabet einer TM. Wir können uns die Berechnung einer konkreten TM (d. h. gegebenes Programm) auch als Abbildung vorstellen:

$$f : \mathcal{A}^* \to \mathcal{A}^*.$$

Hält die TM für einen Eingabewert nicht an, so sei der Wert von f undefiniert.

Dies motiviert folgende allgemeine

Definition: Eine Funktion $f : E \to A$ heisst berechenbar, wenn es einen Algorithmus gibt, der für jede Eingabe $e \in E$, für die $f(e)$ definiert ist, terminiert und das Ergebnis $f(e) \in A$ liefert.

Welche Funktionen sind in diesem Sinne berechenbar?

Auf einem PC mit unendlich viel Speicher könnte man mit Leichtigkeit eine TM simulieren. Das bedeutet, dass man zu jeder TM ein äquivalentes PC-Programm erzeugen kann, welches das Verhalten der TM Schritt für Schritt nachvollzieht. Ein PC (mit unendlich viel Speicher) kann daher alles berechnen, was eine TM berechnen kann.

Interessanter ist aber, dass man zeigen kann, dass die TM trotz ihrer Einfachheit alle Berechnungen durchführen kann, zu denen der PC in der Lage ist. Zu einem PC mit gegebenem Programm kann man also eine TM angeben, die die Berechnung des PCs nachvollzieht! Computer und TM können dieselbe Klasse von Problemen berechnen!

Bemerkung: Im Laufe von Jahrzehnten hat man viele (theoretische und praktische) Berechnungsmodelle erfunden. Die TM ist nur eines davon. Jedes Mal hat sich herausgestellt: Hat eine Maschine gewisse Mindesteigenschaften, so kann sie genausoviel wie eine TM berechnen. Dies nennt man Turing-Äquivalenz.

Die Church'sche[4] These lautet daher:

> Alles was man für intuitiv berechenbar hält kann man mit einer TM ausrechnen.

Dabei heißt intuitiv berechenbar, dass man einen Algorithmus dafür angeben kann.

Mehr dazu in Theoretische Informatik.

Folgerung: Berechenbare Probleme kann man mit fast jeder Computersprache lösen. Unterschiede bestehen aber in der Länge und Eleganz der dafür nötigen Programme, sowie der zur Erstellung notwendigen Zeit (Auch die Effizienz ihrer Ausführung kann sehr unterschiedlich sein, allerdings hängt dieser Punkt sehr von der Compilerimplementation ab.)

Bemerkung: Es gibt auch nicht berechenbare Probleme! So kann man z. B. keine TM angeben, die für jede gegebene TM entscheidet, ob diese den Endzustand

[4] Alonzo Curch, US-amerikanischer Mathematiker, Logiker und Philosoph, 1903–1995

erreicht oder nicht (Halteproblem).

Dieses Problem ist aber noch partiell-berechenbar, d. h. für jede terminierende TM erfährt man dies nach endlicher Zeit, für jede nicht-terminierende TM erfährt man aber kein Ergebnis.

Reale Computer

Algorithmen waren schon vor der Entwicklung unserer heutigen Computer bekannt, allerdings haperte es mit der Ausführung. Zunächst arbeiteten Menschen als „Computer“!

- Lewis Fry Richardson[5] schlägt in seinem Buch *Weather Prediction by Arithmetical Finite Differences* vor, das Wetter für den nächsten Tag mit 64000 (!) menschlichen Computern auszurechnen. Der Vorschlag wird als unpraktikabel verworfen.

- In Los Alamos werden Lochkartenmaschinen und menschliche Rechner für Berechnungen eingesetzt. Richard Feynman[6] organisierte sogar einen Wettbewerb zwischen beiden.

[5] Lewis Fry Richardson, brit. Meteorologe, 1881–1953.

[6] Richard P. Feynman, US-amerik. Physiker, Nobelpreis 1965, 1918–1988.

Menschliche Rechner Ende des 19. Jahrhunderts

Der Startpunkt der Entwicklung realer Computer stimmt (zufällig?) relativ genau mit der Entwicklung theoretischer Berechenbarkeitskonzepte durch Church und Turing überein.

Dabei verstehen wir Computer bzw. (Universal-)Rechner als Maschinen zur Ausführung beliebiger Algorithmen in obigem Sinne (d. h. sie können nicht „nur“ rechnen im Sinne arithmetischer Operationen).

Einige der wichtigsten frühen Rechenmaschinen waren:

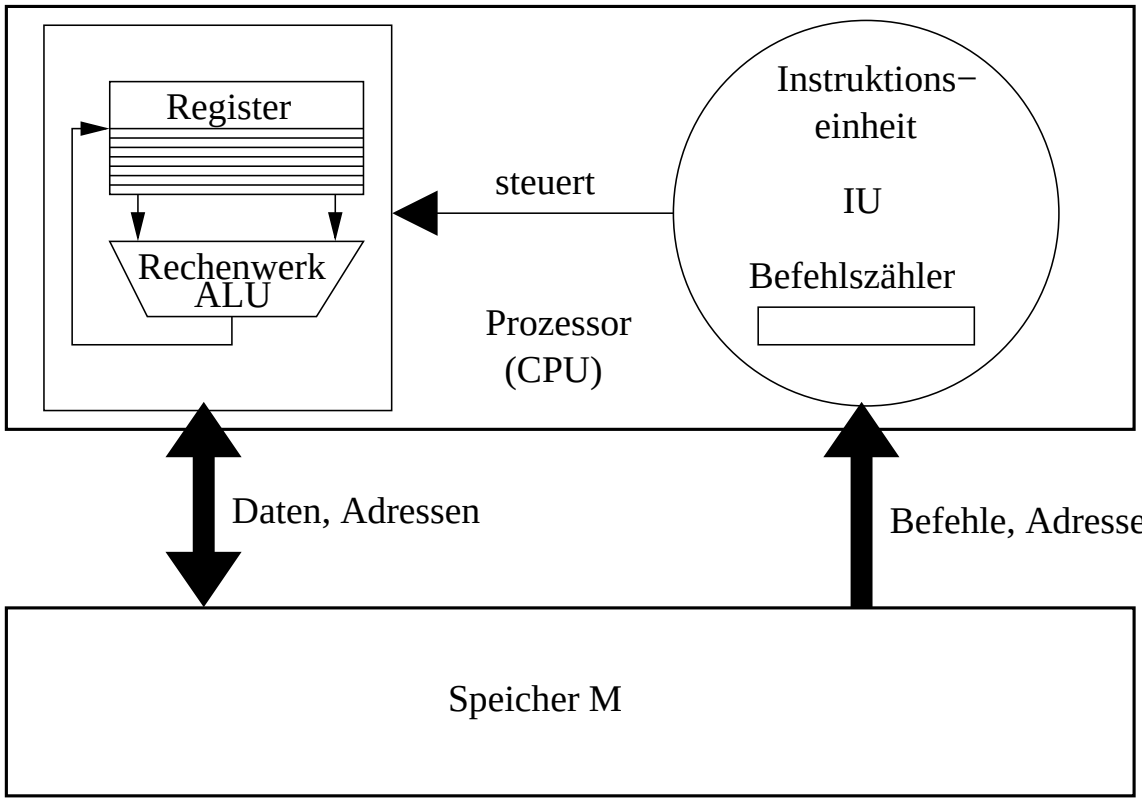
- Zuse Z3, Mai 1941, mechanisch, turing-vollständig (aber nicht als solcher konstruiert), binäre Gleitkommaarithmetik
- Atanasoff-Berry-Computer, Sommer 1941, elektronisch (Röhren), nicht turing-mächtig, gebaut zur Lösung linearer Gleichungssysteme (29×29)
- Colossus, 1943, elektronisch, nicht turing-mächtig, Kryptographie

- Mark 1, 1944, mechanisch, turing-vollständig, Ballisitik
- ENIAC, 1946, elektronisch, turing-vollständig, Ballistik
- EDVAC, 1949, elektronisch, turing-vollständig, Ballistik, erste „Von-Neumann-Architektur“

Praktische Computer basieren meist auf dem von John von Neumann 1945 im Rahmen der EDVAC-Entwicklung eingeführten Konzept. Es ist umstritten welche der Ideen tatsächlich genau von ihm sind.

Geschichte: John von Neumann[7] war einer der bedeutendsten Mathematiker. Von ihm stammt die Spieltheorie, die mathematische Begründung der Quantenmechanik, sowie wichtige Beiträge zu Informatik und Numerik.

[7] János Neumann Margittai, Mathematiker österreichisch-ungarischer Herkunft, 1903–1957.



Der Speicher M besteht aus endlich vielen Feldern, von denen jedes eine Zahl aufnehmen kann. Im Unterschied zur TM kann auf jedes Feld ohne vorherige Positionierung zugegriffen werden (wahlfreier Zugriff, random access).

Zum Zugriff auf den Speicher wird ein Index, auch Adresse genannt, verwendet, d. h. wir können den Speicher als Abbildung

$$M : A \to D$$

auffassen.

Für die Adressen gilt $A = [0, N-1] \subset \mathbb{N}_0$ wobei aufgrund der binären Organisation $N = 2^n$ gilt. n ist die Anzahl der erforderlichen Adressleitungen.

Für D gilt $D = [0, 2^m - 1]$ mit der Wortbreite m, die meistens ein Vielfaches von 8 ist. m ist die Anzahl der erforderlichen Datenleitungen.

Die Gesamtkapazität des Speichers ist demnach $m \cdot 2^n$ Bit. Jedes Bit kann zwei Werte annehmen, 0 oder 1. In der Praxis wird die Größe des Speichers in Byte angegeben, wobei ein Byte aus 8 Bit besteht. Damit enthält ein Speicher mit n Adressleitungen bei Wortbreite m genau $(m/8) \cdot 2^n$ Byte.

Gebräuchlich sind auch noch die Abkürzungen 1 Kilobyte $= 2^{10}$ Byte $= 1024$ Byte,

1 Megabyte $= 2^{20}$ Byte, 1 Gigabyte $= 2^{30}$ Byte.

Der Speicher enthält sowohl Daten (das Band in der TM) als auch Programm (die Tabelle in der TM). Den einzelnen Zeilen der Programmtabelle der TM entsprechen beim von Neumannschen Rechner die Befehle. Die Vereinigung von Daten und Programm im Speicher (stored program computer) war der wesentliche Unterschied zu den früheren Ansätzen.

Befehle werden von der Instruktionseinheit (instruction unit, IU) gelesen und dekodiert.

Die Instruktionseinheit steuert das Rechenwerk, welches noch zusätzliche Daten aus dem Speicher liest bzw. Ergebnisse zurückschreibt.

Die Maschine arbeitet zyklisch die folgenden Aktionen ab:

- Befehl holen
- Befehl dekodieren
- Befehl ausführen

Dies nennt man Befehlszyklus. Viel mehr über Rechnerhardware erfährt man in der Vorlesung „Technische Informatik“.

Bemerkung: Hier wurde insbesondere die Interaktion von Rechnern mit der Umwelt, die sog. Ein- und Ausgabe, in der Betrachtung vernachlässigt. Moderne Rechner haben insbesondere die Fähigkeit, auf äußere Einwirkungen hin (etwa Tastendruck) den Programmfluss zu unterbrechen und an anderer Stelle (Turingmaschine: in anderem Zustand) wieder aufzunehmen. Von Neumann hat die Ein-/Ausgabe im Design des EDVAC schon ausführlich beschrieben.

Bemerkung: Heutige Rechner sind wesentlich komplizierter als dieses einfache Modell. Insbesondere sind viele Möglichkeiten der **parallelen** Verarbeitung enthalten. Wichtige Konzepte in modernen Rechnern sind:

- Hierarchisch organisierter Speicher mit Caches
- Pipelining des Befehlsholzyklus
- SIMD Instruktionen, Superskalarität
- Multicorerechner

Programmiersprachen

Die Befehle, die der Prozessor ausführt, nennt man Maschinenbefehle oder auch Maschinensprache. Sie ist relativ umständlich, und es ist sehr mühsam größere Programme darin zu schreiben. Andererseits können ausgefeilte Programme sehr kompakt sein und sehr effizient ausgeführt werden.

Beispiel: Ein Schachprogramm auf einem 6502-Prozessor findet man unter

`http://www.6502.org/source/games/uchess/uchess.pdf`

Es benötigt weniger als 1KB an Speicher!

Die weitaus meisten Programme werden heute in sogenannten höheren Programmiersprachen erstellt. Sinn einer solchen Sprache ist, dass der Programmierer Programme möglichst

- schnell (in Sinne benötigter Programmiererzeit) und
- korrekt (Programm löst Problem korrekt)

erstellen kann.

Wir lernen in dieser Vorlesung die Sprache C++. C++ ist eine Weiterentwicklung der Sprache C, die Ende der 1960er Jahre entwickelt wurde.

Warum C++ ?

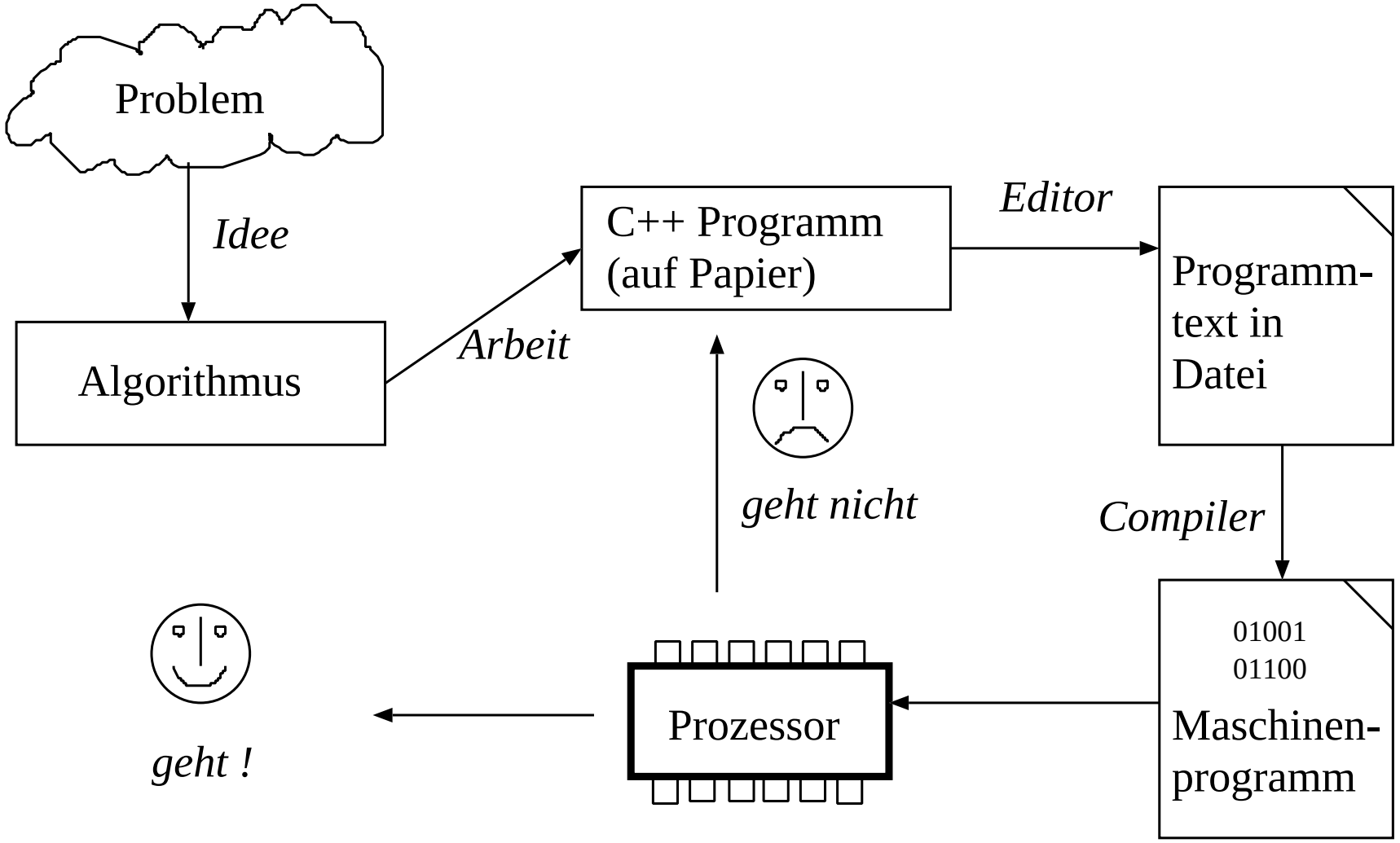
- C++ ist ein ausgereifte Sprache, die für sehr viele große Softwareprojekte verwendet und ständig weiterentwickelt wird.
- C++ unterstützt viele verschiedene Programmierstile: funktional (eingeschränkt), prozedural, objektorientiert, generisch.
- Objektorientierung: Modellierung komplexer Daten ist wichtiger als der Kontrollfluss.
- C++ erlaubt sowohl maschinennahe Programmierung als auch hohes Abstraktionsniveau.
- C++ ist sehr komplex, daher werden wir uns auf einen Ausschnitt beschränken. Vollständigkeit ist kein Ziel dieser Vorlesung!

Programme in einer Hochsprache lassen sich *automatisch* in Programme der Maschinensprache übersetzen. Ein Programm, das dies tut, nennt man Übersetzer oder Compiler.

Ein Vorteil dieses Vorgehens ist auch, dass Programme der Hochsprache in verschiedene Maschinensprachen (Portabilität) übersetzt und andererseits verschiedene Hochsprachen auch in ein und dieselbe Maschinensprache übersetzt werden können (Flexibilität).

Es gibt auch sog. interpretierte Sprachen. Dort werden die Anweisungen der Hochsprache während der Ausführung ,,on the fly" in Maschinensprache übersetzt. Beispiele: Python, Shell, Basic.

Schließlich gibt es Mischformen, bei denen von der Hochsprache in eine Zwischensprache übersetzt wird, die dann interpretiert wird. Beispiel: Java.



Die einzelnen Schritte bei der Programmerstellung im Überblick.

Warum gibt es verschiedene Programmiersprachen ?

Wie bei der Umgangssprache: teils sind Unterschiede historisch gewachsen, teils sind die Sprachen wie Fachsprachen auf verschiedene Problemstellungen hin optimiert.

Andererseits sind die Grundkonzepte in vielen prozeduralen bzw. objektorientierten Sprachen sehr ähnlich. Eine neue Sprache in dieser Klasse kann relativ leicht erlernt werden.

Komplexität von Programmen

Die Leistungsfähigkeit von Computern wächst schnell.

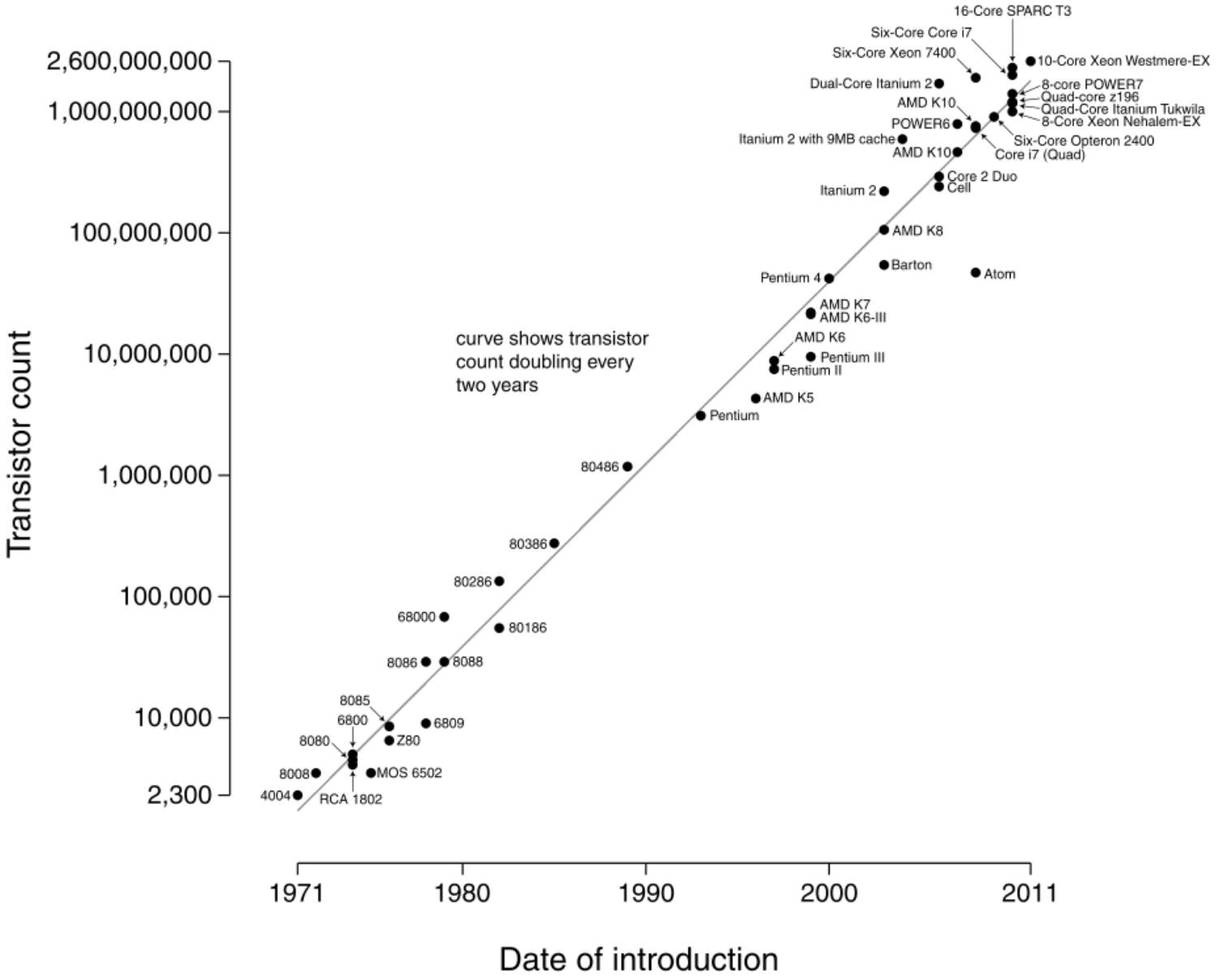
Wissen: (Moore'sches[8] „Gesetz“)
Die Anzahl der Transistoren pro Flächeneinheit auf einem Halbleiterchip verdoppelt sich etwa alle 18–24 Monate.

Folgende Grafik (Wikipedia, verfügbar unter Creative Commons-Lizenz, Urheber Wgsimon) zeigt die Anzahl der Transistoren verschiedener Halbleiterchips und ihr Einführungsdatum.

Stand 2015: Bis zu 8 Mrd. Transistoren (NVidia Maxwell GM200).

[8] Gordon E. Moore, US-amerk. Unternehmer (Mitbegründer der F. Intel), geb. 1929.

Microprocessor Transistor Counts 1971-2011 & Moore's Law



Beispiel: Entwicklung von Taktgeschwindigkeit, Speichergröße und Größe des Linux-Kernel.

Zeit	Proz	Takt	RAM	Disk	Linux Kernel (.tar.gz)
1982	Z80	6	64KB	800KB	6KB (CPM)
1988	80286	10	1MB	20MB	20KB (DOS)
1992	80486	25	20MB	160MB	140KB (0.95)
1995	PII	100	128MB	2GB	2.4MB (1.3.0)
1999	PII	400	512MB	10GB	13.2MB (2.3.0)
2001	PIII	850	512MB	32GB	23.2MB (2.4.0)
2004	P4 (Prescott)	3.8 GHz	2048 MB	250 GB	36 MB (2.4.26)
2010	i7 (Westmere)	3.5 GHz	8196 MB	1024 GB	84 MB (2.6.37.7)

Bis 2001 exponentielles Wachstum. Prozessortaktfrequenz stagniert seit 2004. Wachstum des Linux-Kernel ist auch abgeflacht.

Grenzen der Hardwareentwicklung:

- Je höher die Takfrquenz desto höher die Wärmeentwicklung.
- Mehr an Transistoren wird in viele unabhängige Cores gesteckt.
- Erfordert parallele Programmierung.

Grenzen der Softwareentwicklung:

- Die benötigte Zeit zum Erstellen großer Programme skaliert mehr als linear, d. h. zum Erstellen eines doppelt so großen Programmes braucht man mehr als doppelt so lange.
- Verbesserte Programmiertechnik, Sprachen und Softwareentwurfsprozesse. Einen wesentlichen Beitrag leistet hier die objektorientierte Programmierung, die wir in dieser Vorlesung am Beispiel von C++ erlernen werden.

Funktionale Programmierung

Wichtig:

- In diesem Abschnitt beschränken wir uns bewusst auf eine sehr kleine Teilmenge von C++.
- In den Übungen sind nur die vorgestellten Befehle erlaubt!
- Abgaben die insbesondere Schleifen, Variablen und Zuweisung verwenden werden mit Null Punkten bewertet!
- Wenn Sie noch nicht wissen was das ist: Umso besser!

Auswertung von Ausdrücken

Arithmetische Ausdrücke

Beispiel: Auswertung von:

$$5 + 3 \text{ oder } ((3 + (5 * 8)) - (16 * (7 + 9))).$$

Programm: (Erste Schritte [erstes.cc])

```
#include "fcpp.hh"

int main()
{
  return print( (3+(5*8)) - (16*(7+9)) );
}
```

Übersetzen (in Unix-Shell):

```
> g++ -o erstes erstes.cc
```

Ausführung:

```
> ./erstes
-213
```

Bemerkung:

- Ohne „-o erstes“ wäre der Name „a.out“ verwendet worden.
- Das Programm berechnet den Wert des Ausdrucks und druckt ihn auf der Konsole aus.

Wie wertet der Rechner so einen Ausdruck aus?

Die Auswertung eines zusammengesetzten Ausdruckes lässt sich auf die Auswertung der vier elementaren Rechenoperationen $+, -, *$ und $/$ zurückführen.

Dazu fassen wir die Grundoperationen als zweistellige Funktionen auf:

$$+, -, *, / \ : \ \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}.$$

Jeden Ausdruck können wir dann äquivalent umformen:

$$((3 + (5 * 8)) - (16 * (7 + 9))) \ \equiv \ -(+(3, *(5, 8)), *(16, +(7, 9))).$$

Definition: Die linke Schreibweise nennt man Infix-Schreibweise (*infix notation*), die rechte Präfix-Schreibweise (*prefix notation*).

Bemerkung: Die Infix-Schreibweise ist für arithmetische Ausdrücke bei Hinzunahme von Präzedenzregeln wie „Punkt vor Strich“ und dem Ausnutzen des Assoziativgesetzes kürzer (da Klammern wegelassen werden können) und leichter lesbar als die Präfix-Schreibweise.

Bemerkung: Es gibt auch eine Postfix-Schreibweise, welche zum Beispiel in HP-Taschenrechnern, dem Emacs-Programm „Calc“ oder der Computersprache Forth verwendet wird.

Die vier Grundoperationen $+, -, *, /$ betrachten wir als atomar. Im Rechner gibt es entsprechende Baugruppen, die diese atomaren Operationen realisieren.

Der Compiler übersetzt den Ausdruck aus der Infix-Schreibweise in die äquivalente Präfixschreibweise. Die Auswertung des Ausdrucks, d. h. die Berechnung der Funktionen, erfolgt dann von innen nach aussen:

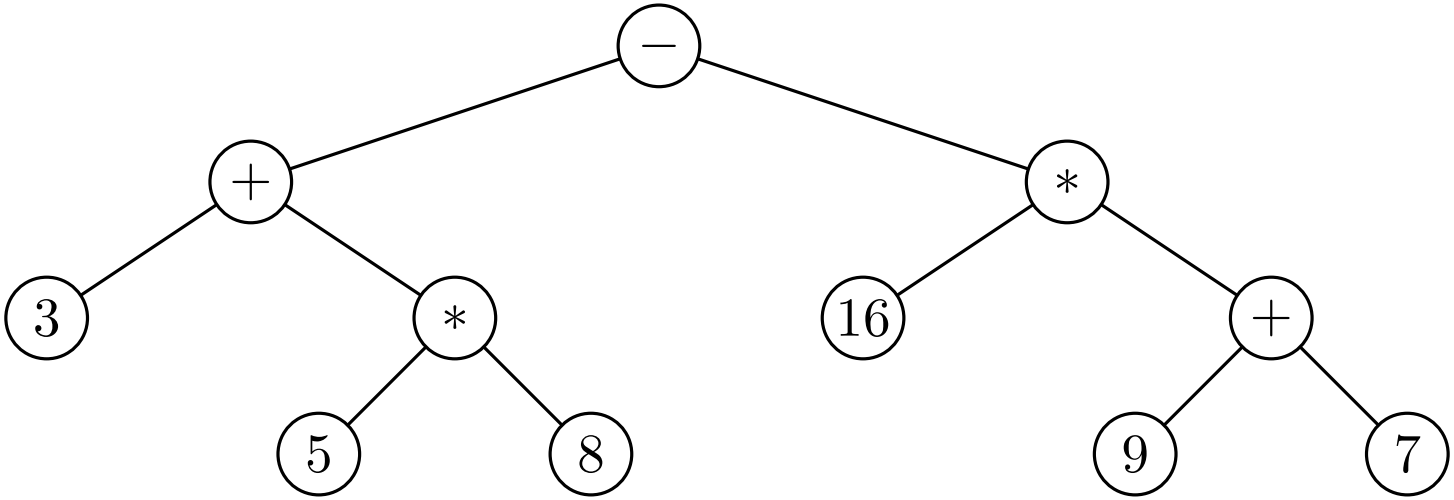
```
  -(+(3,*(5,8)),*(16,+(7,9)))
= -(+(3,  40  ),*(16,+(7,9)))
= -(    43     ,*(16,+(7,9)))
= -(    43     ,*(16,  16  ))
= -(    43     ,    256     )
=            -213
```

Bemerkung: Dies ist nicht die einzig mögliche Reihenfolge der Auswertung der Teiloperationen, alle Reihenfolgen führen jedoch zum gleichen Ergebnis! (Zumindest bei nicht zu großen, ganzen Zahlen).

Bemerkung: C++ kennt die Punkt-vor-Strich-Regel und das Assoziativgesetz. Überflüssige Klammern können also weggelassen werden.

Ausdrücke als Bäume

Jeder arithmetische Ausdruck kann als binärer Baum dargestellt werden. Die Auswertung des Ausdruckes erfolgt dann von den Blättern zur Wurzel. In dieser Darstellung erkennt man welche Ausführungsreihenfolgen möglich sind bzw. welche Teilausdrück gleichzeitig ausgewertet werden können (Datenflussgraph).



Funktionen

Zu den schon eingebauten Funktionen wie $+, -, *, /$ kann man noch weitere benutzerdefinierte Funktionen hinzuzufügen.

Beispiel: Eine einstellige Funktion:

```
int quadrat( int x )
{
  return x*x;
}
```

Die erste Zeile (Funktionskopf) vereinbart, dass die neue Funktion namens `quadrat` als Argument eine Zahl mit Namen `x` vom Typ `int` als Eingabe bekommt und einen Wert vom Typ `int` als Ergebnis liefert.

Der Funktionsrumpf (*body*) zwischen geschweiften Klammern sagt, was die Funktion tut. Der Ausdruck nach return ist der Rückgabewert.

Wir werden uns zunächst auf einen sehr kleinen Teil des Sprachumfangs von C/C++ beschränken. Dort besteht der Funktionsrumpf nur aus dem Wort return gefolgt von einem Ausdruck gefolgt von einem Semikolon.

Bemerkung: C++ ist eine streng typgebundene Programmiersprache (*strongly typed*), d. h. jedem Bezeichner (z. B. `x` oder `quadrat`) ist ein Typ zugeordnet. Diese Typzuordnung kann nicht geändert werden (statische Typbindung, *static typing*).

Dies ist in völliger Analogie zur Mathematik:

$$x \in \mathbb{Z}, \qquad f : \mathbb{N} \to \mathbb{N}.$$

Bemerkung: Der Typ `int` entspricht dabei (kleinen) ganzen Zahlen. Andere Typen sind `float`, `double`, `char`, `bool`. Später werden wir sehen, dass man auch neue Typen hinzufügen kann.

Programm: (Verwendung [quadrat.cc])

```
#include "fcpp.hh"

int quadrat( int x )
{
  return x*x;
}

int main()
{
  return print( quadrat( 3 ) + quadrat( 4+4 ) );
}
```

Bemerkung: Damit können wir die Bedeutung aller Elemente des Programmes verstehen.

- Neue Funktionen kann man (in C) nur in Präfix-Schreibweise verwenden.
- `main` ist eine Funktion ohne Argumente und mit Rückgabetyp `int`.
- #include "fcpp.hh" ist ein sogenannter Include-Befehl. Er sorgt dafür, dass die in der Datei fcpp.hh enthaltenen Erweiterungen von C++, etwa zusätzliche Funktionen, verwendet werden können. fcpp.hh ist nicht Teil des C++ Systems, sondern wird von uns für die Vorlesung zur Verfügung gestellt (erhältlich auf der Webseite). Achtung: Die Datei muss sich im selben Verzeichnis befinden wie das zu übersetzende Programm damit der Compiler diese finden kann.
- `print` ist eine Funktion mit Rückgabewert 0 (unabhängig vom Argument), welche den Wert des Arguments auf der Konsole ausdruckt (Seiteneffekt). Die Definition dieser Funktion ist in der Datei `fcpp.hh` enthalten.

- Die Programmausführung beginnt immer mit der Funktion `main` (sozusagen das Startsymbol).

Selektion

Fehlt noch: Steuerung des Programmverlaufs in Abhängigkeit von Daten.

Beispiel: Betragsfunktion

$$|x| = \begin{cases} -x & x < 0 \\ x & x \geq 0 \end{cases}$$

Um dies ausdrücken zu können, führen wir eine spezielle dreistellige Funktion `cond` ein:

Programm: **(Absolutwert [absolut.cc])**

```
#include "fcpp.hh"

int absolut( int x )
{
  return cond( x<=0, -x, x );
}

int main()
{
  return print( absolut( -3 ) );
}
```

Der Operator `cond` erhält drei Argumente: Einen Boolschen Ausdruck und zwei normale Ausdrücke. Ein Boolscher Ausdruck hat einen der beiden Werte „wahr“ oder „falsch“ als Ergebnis. Ist der Wert „wahr“, so ist das Resultat des `cond`-Operators der Wert des zweiten Arguments, ansonsten der des dritten.

Bemerkung: `cond` kann keine einfache Funktion sein:

- `cond` kann auf verschiedene Typen angewendet werden, und auch der Typ des Rückgabewerts steht nicht fest.
- Oft wird `cond` nicht alle Argumente auswerten dürfen, um nicht in Fehler oder Endlosschleifen zu geraten.

Bemerkung: Damit haben wir bereits eine Menge von Konstrukten kennengelernt, die turing-äquivalent ist!

Programm: **(Elementare funktionale Programmierung [alles_funktional.cc])**

```
#include "fcpp.hh"

int quadrat( int x ) { return x*x; }

int absolut( int x ) { return cond( x<=0, -x, x ); }

int main()
{
  return print( absolut(-4) * (7*quadrat(3)+8) );
}
```

Syntaxbeschreibung mit Backus-Naur Form

EBNF

Die Regeln nach denen wohlgeformte Sätze einer Sprache erzeugt werden, nennt man Syntax.

Die Syntax von Programmiersprachen ist recht einfach. Zur Definition verwendet man eine spezielle Schreibweise, die erweiterte Backus[9]-Naur[10] Form (EBNF):

Man unterscheidet in der EBNF folgende Zeichen bzw. Zeichenketten:

- Unterstrichene Zeichen oder Zeichenketten sind Teil der zu bildenden, wohlgeformten Zeichenkette. Sie werden nicht mehr durch andere Zeichen ersetzt, deshalb nennt man sie terminale Zeichen.

- Zeichenketten in spitzen Klammern, wie etwa $<$ Z $>$ oder $<$ Ausdruck $>$ oder $<$ Zahl $>$, sind Symbole für noch zu bildende Zeichenketten. Regeln beschrei-

[9] John Backus, 1924–2007, US-amerik. Informatiker.

[10] Peter Naur, geb. 1928, dänischer Informatiker.

ben, wie diese Symbole durch weitere Symbole und/oder terminale Zeichen ersetzt werden können. Da diese Symbole immer ersetzt werden, nennt man sie nichtterminale Symbole.

- $< \epsilon >$ bezeichnet das „leere Zeichen“.

- Die normal gesetzten Zeichen(ketten)

  $$::= \quad | \quad \{ \quad \} \quad \}^+ \quad [\quad]$$

  sind Teil der Regelbeschreibung und tauchen nie in abgeleiteten Zeichenketten auf. (Es sei denn sie sind unterstrichen und somit terminale Zeichen).

- (Alternativ findet man auch die Konvention terminale Symbole in Anführungszeichen zu setzen und die spitzen Klammern bei nichtterminalen wegzulassen).

Jede Regel hat ein Symbol auf der linken Seite gefolgt von „::=“. Die rechte Seite beschreibt, durch was das Symbol der linken Seite ersetzt werden kann.

Beispiel:

$$<\text{A}> \quad ::= \quad \underline{\text{a}} <\text{A}> \underline{\text{b}}$$
$$<\text{A}> \quad ::= \quad <\epsilon>$$

Ausgehend vom Symbol $<\text{A}>$ kann man somit folgende Zeichenketten erzeugen:

$$<\text{A}> \rightarrow \underline{\text{a}}<\text{A}>\underline{\text{b}} \rightarrow \underline{\text{aa}}<\text{A}>\underline{\text{bb}} \rightarrow \ldots \rightarrow \underbrace{\underline{a}\ldots\underline{a}}_{n \text{ mal}}<\text{A}>\underbrace{\underline{b}\ldots\underline{b}}_{n \text{ mal}} \rightarrow \underbrace{\underline{a}\ldots\underline{a}\underline{b}\ldots\underline{b}}_{n \text{ mal} \quad n \text{ mal}}$$

Bemerkung: Offensichtlich kann es für ein Symbol mehrere Ersetzungsregeln geben. Wie im MIU-System ergeben sich die wohlgeformten Zeichenketten durch alle möglichen Regelanwendungen.

Kurzschreibweisen

Oder:

Das Zeichen „ | “ („oder“) erlaubt die Zusammenfassung mehrerer Regeln in einer Zeile. Beispiel: $<\text{A}> ::= \underline{\text{a}} <\text{A}> \underline{\text{b}} \mid <\epsilon>$

Option:

$<\text{A}> ::= [<\text{B}>]$ ist identisch zu $<\text{A}> ::= <\text{B}> \mid <\epsilon>$

Wiederholung mit $n \geq 0$:

$<\text{A}> ::= \{ <\text{B}> \}$ ist identisch mit $<\text{A}> ::= <\text{A}> <\text{B}> \mid <\epsilon>$

Wiederholung mit $n \geq 1$:

$<\text{A}> ::= \{ <\text{B}> \}^+$ ist identisch zu
$<\text{A}> ::= <\text{A}> <\text{B}> \mid <\text{B}>$

Beispiel: Die EBNF wird auch in der UNIX Dokumentation verwendet:

```
BSDTAR(1)  BSD General Commands Manual        BSDTAR(1)

NAME
     tar -- manipulate tape archives

SYNOPSIS
     tar [bundled-flags <args>] [<file> | <pattern> ...]
     tar {-c} [options] [files | directories]
     tar {-r | -u} -f archive-file [options] [files | directories]
     tar {-t | -x} [options] [patterns]
```

Syntaxbeschreibung für FC++

Die bisher behandelte Teilmenge von C++ nennen wir FC++ („funktionales C++“) und wollen die Syntax in EBNF beschreiben.

Syntax: (Zahl)

$$<\text{Zahl}> ::= [\ \underline{+} \mid \underline{-}\ ]\ \{\ <\text{Ziffer}>\ \}^{+}$$

Syntax: (Ausdruck)

$$\begin{array}{lll}
<\text{Ausdruck}> & ::= & <\text{Zahl}> \mid [\ \underline{-}\ ] <\text{Bezeichner}> \mid \\
 & & \underline{(}\ <\text{Ausdruck}> <\text{Operator}> <\text{Ausdruck}>\ \underline{)} \mid \\
 & & <\text{Bezeichner}>\ \underline{(}\ [\ <\text{Ausdruck}>\ \{\ \underline{,}\ <\text{Ausdruck}>\ \}\ ]\ \underline{)} \mid \\
 & & <\text{Cond}> \\
<\text{Bezeichner}> & ::= & <\text{Buchstabe}>\ \{\ <\text{Buchstabe oder Zahl}>\ \} \\
<\text{Operator}> & ::= & \underline{+} \mid \underline{-} \mid \underline{*} \mid \underline{/}
\end{array}$$

Weggelassen: Regeln für $<\text{Buchstabe}>$ und $<\text{Buchstabe oder Zahl}>$.

Diese einfache Definition für Ausdrücke enthält weder Punkt-vor-Strich noch das Weglassen von Klammern aufgrund des Assoziativgesetzes!

Hier die Syntax einer Funktionsdefinition in EBNF:

Syntax: (Funktionsdefinition)

<Funktion>	::=	<Typ> <Name> (<formale Parameter>) { <Funktionsrumpf> }
<Typ>	::=	<Bezeichner>
<Name>	::=	<Bezeichner>
<formale Parameter>	::=	[<Typ> <Name> { , <Typ> <Name> }]

Die Argumente einer Funktion in der Funktionsdefinition heissen formale Parameter. Sie bestehen aus einer kommaseparierten Liste von Paaren aus Typ und Name. Damit kann man also n-stellige Funktionen mit $n \geq 0$ erzeugen.

Regel für den Funktionsrumpf:

<Funktionsrumpf> ::= return <Ausdruck> ;

Hier ist noch die Syntax für die Selektion:

Syntax: **(Cond)**

<Cond>	::=	cond (<BoolAusdr> , <Ausdruck> , <Ausdruck>)
<BoolAusdr>	::=	true | false | (<Ausdruck> <VglOp> <Ausdruck>) |
		(<BoolAusdr> <LogOp> <BoolAusdr>) |
		! (<BoolAusdr>)
<VglOp>	::=	== | != | < | > | <= | >=
<LogOp>	::=	&& | ||

Bemerkung: Beachte dass der Test auf Gleichheit als == geschrieben wird!

Syntax: (FC++ Programm)

<FC++-Programm>	::=	{ <Include> } { <Funktion> }$^+$
<Include>	::=	#include " <DateiName> "

Bemerkung: **(Leerzeichen)** C++ Programme erlauben das Einfügen von Leerzeichen, Zeilenvorschüben und Tabulatoren („whitespace“) um Programme für den Menschen lesbarer zu gestalten. Hierbei gilt folgendes zu beachten:

- Bezeichner, Zahlen, Schlüsselwörter und Operatorzeichen dürfen keinen Whitespace enthalten:
  - `zaehler` statt `zae hler`,
  - `893371` statt `89 3371`,
  - `return` statt `re tur n`,
  - `&&` statt `& &`.

- Folgen zwei Bezeichner, Zahlen oder Schlüsselwörter nacheinander so muss ein Whitespace (also mindestens ein Leerzeichen) dazwischen stehen:
  - `int f(int x)` statt `intf(intx)`,
  - `return x;` statt `returnx;`.

Die obige Syntaxbeschreibung mit EBNF ist nicht mächtig genug, um fehlerfrei übersetzbare C++ Programme zu charakterisieren. So enthält die Syntaxbeschreibung üblicherweise nicht solche Regeln wie:

- Kein Funktionsname darf doppelt vorkommen.

- Genau eine Funktion muss `main` heissen.

- Namen müssen an der Stelle bekannt sein wo sie vorkommen.

Bemerkung: Mit Hilfe der EBNF lassen sich sogenannte kontextfreie Sprachen definieren. Entscheidend ist, dass in EBNF-Regeln links immer nur genau ein nichtterminales Symbol steht. Zu jeder kontextfreien Sprache kann man ein Programm (genauer: einen Kellerautomaten) angeben, das für jedes vorgelegte Wort in endlicher Zeit entscheidet, ob es in der Sprache ist oder nicht. Man sagt: kontextfreie Sprachen sind entscheidbar. Die Regel „Kein Funktionsname darf doppelt vorkommen“ lässt sich mit einer kontextfreien Sprache nicht formulieren und wird deshalb extra gestellt.

Kommentare

Mit Hilfe von Kommentaren kann man in einem Programmtext Hinweise an einen menschlichen Leser einbauen. Hier bietet C++ zwei Möglichkeiten an:

```
// nach // wird der Rest der Zeile ignoriert
/* Alles dazwischen ist Kommentar  (auch über
    mehrere Zeilen)
*/
```

Vorlesung heute:

- Syntax und Semantik, Substitutionsmodell.
- Linear rekursiver Prozess
- Linear iterativer Prozess
- Baumrekursiver Prozess

Das Substitutionsmodell

Selbst wenn ein Programm vom Übersetzer fehlerfrei übersetzt wird, muss es noch lange nicht korrekt funktionieren. Was das Programm tut bezeichnet man als $Semantik$ (Bedeutungslehre). Das in diesem Abschnitt vorgestellte Substitutionsmodell kann die Wirkungsweise funktionaler Programme beschreiben.

Definition: (Substitutionsmodell) Die Auswertung von Ausdrücken geschieht wie folgt:

1. <Zahl> wird als die Zahl selbst ausgewertet.

2. < Name > ($<a_1>$, $<a_2>$, . . . , $<a_n>$) wird für Elementarfunktionen folgendermaßen ausgewertet:

   (a) Werte die Argumente aus. Diese sind wieder Ausdrücke. Unsere Definition ist also rekursiv!
   (b) Werte die Elementarfunktion <Name> auf den so berechneten Werten aus.

3. $<$Name$>$ ($<a_1>$, $<a_2>$, . . . , $<a_n>$) wird für benutzerdefinierte Funktionen folgendermaßen ausgewertet:

   (a) Werte die Argumente aus.
   (b) Werte den Rumpf der Funktion $<$Name$>$ aus, wobei jedes Vorkommen eines formalen Parameters durch den entsprechenden Wert des Arguments ersetzt wird. Der Rumpf besteht im wesentlichen ebenfalls wieder aus der Auswertung eines Ausdrucks.

4. cond ($<a_1>$, $<a_2>$, $<a_3>$) wird ausgewertet gemäß:

   (a) Werte $<a_1>$ aus.
   (b) Ist der erhaltene Wert `true`, so erhält man den Wert des cond-Ausdrucks durch Auswertung von $<a_2>$, ansonsten von $<a_3>$. Wichtig: nur eines der beiden Argumente $<a_2>$ oder $<a_3>$ wird ausgewertet.

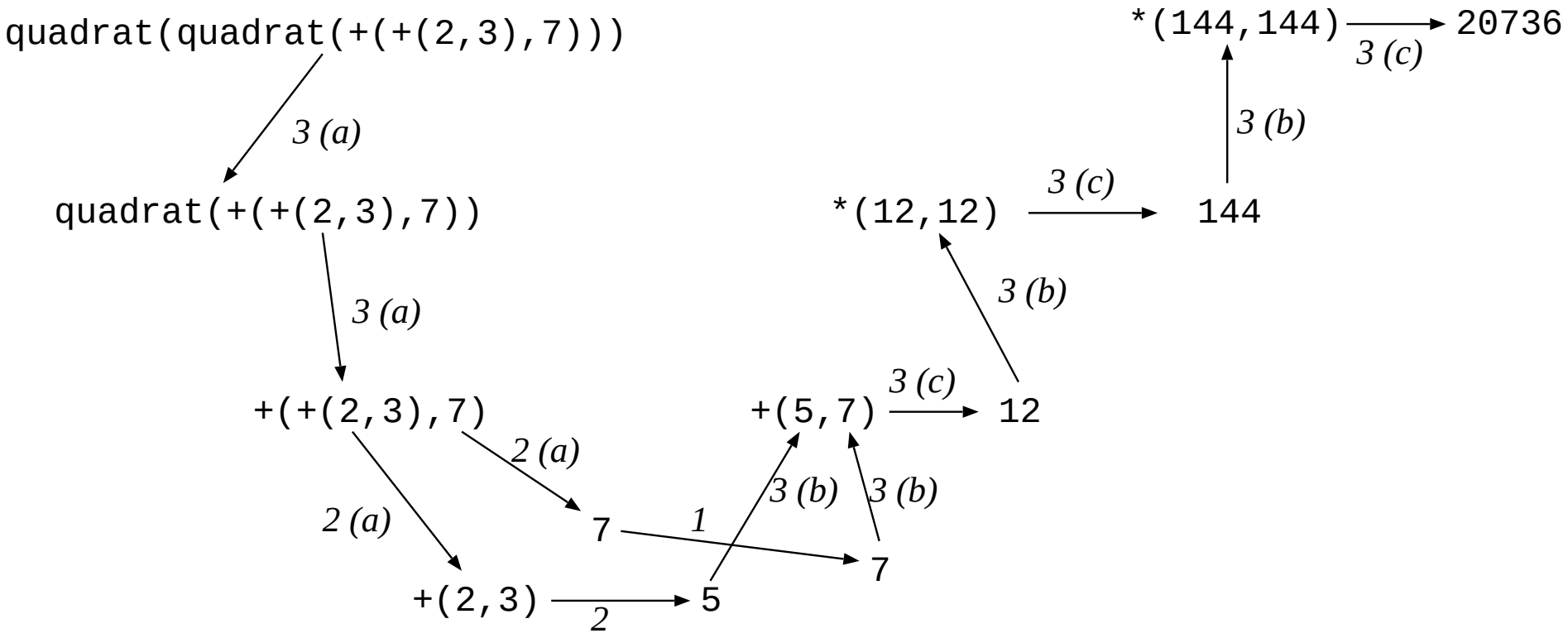
Bemerkung: Die Namen der formalen Parameter sind egal, sie entsprechen sogenannten gebundenen Variablen in logischen Ausdrücken.

Beispiel:

```
quadrat(3) = *( 3, 3 ) = 9
```

Beispiel:

```
quadrat( quadrat( (2+3)+7 ) )
   = quadrat( quadrat( +( +( 2, 3 ), 7 ) ) )
   = quadrat( quadrat( +( 5        , 7 ) ) )
   = quadrat( quadrat( 12                ) )
   = quadrat( *( 12, 12                  ) )
   = quadrat( 144                          )
   = *( 144, 144 )
   = 20736
```



Aufbau einer Kette von verzögerten Funktionsaufrufen.

Linear-rekursive Prozesse

Beispiel: (Fakultätsfunktion) Sei $n \in \mathbb{N}$. Dann gilt

$$\begin{aligned} n! &= \prod_{i=1}^{n} i, \\ &= 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n. \end{aligned}$$

Oder rekursiv:

$$n! = \begin{cases} 1 & n = 1, \\ n(n-1)! & n > 1. \end{cases}$$

Programm: **(Rekursive Berechnung der Fakultät [fakultaet.cc])**

```
#include "fcpp.hh"

int fakultaet( int n )
{
  return cond( n<=1, 1, n*fakultaet(n-1) );
}

int main()
{
  return print( fakultaet(5) );
}
```

Die Auswertung kann mithilfe des Substitutionsmodells wie folgt geschehen:

```
fakultaet(5) = *( 5, fakultaet(4) )
             = *( 5, *( 4, fakultaet(3) ) )
             = *( 5, *( 4, *( 3, fakultaet(2) ) ) )
             = *( 5, *( 4, *( 3, *( 2, fakultaet(1) ) ) ) )
             = *( 5, *( 4, *( 3, *( 2, 1            ) ) ) )
             = *( 5, *( 4, *( 3, 2                  ) ) )
             = *( 5, *( 4, 6                        ) )
             = *( 5, 24                             )
             = 120
```

Definition: Dies bezeichnen wir als linear rekursiven Prozess (die Zahl der verzögerten Operationen wächst linear in n). Die Aufrufe formen eine lineare Kette von Funktionsaufrufen.

Linear-iterative Prozesse

Interessanterweise lässt sich die Kette verzögerter Operationen bei der Fakultätsberechnung vermeiden. Betrachte dazu folgendes Tableau von Werten von n und $n!$:

$$\begin{array}{cccccccccccc} n & 1 & & 2 & & 3 & & 4 & & 5 & & 6 & \ldots \\ & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\ n! & 1 & \to & 2 & \to & 6 & \to & 24 & \to & 120 & \to & 720 & \ldots \end{array}$$

Idee: Führe das Produkt als zusätzliches Argument mit.

Programm: (Iterative Fakultätsberechnung [fakultaetiter.cc])

```
#include "fcpp.hh"

int faklter( int produkt, int zaehler, int ende )
{
  return cond( zaehler>ende,
               produkt,
               faklter( produkt*zaehler, zaehler+1, ende ) );
}

int fakultaet( int n )
{
  return faklter( 1, 1, n );
}

int main()
{
  return print( fakultaet(5) );
}
```

Die Analyse mit Hilfe des Substitutionsprinzips liefert:

```
fakultaet(5) = fakIter( 1, 1, 5 )
             = fakIter( 1, 2, 5 )
             = fakIter( 2, 3, 5 )
             = fakIter( 6, 4, 5 )
             = fakIter( 24, 5, 5 )
             = fakIter( 120, 6, 5 )
             = 120
```

Hier wird allerdings von folgender Optimierung ausgegangen: In fakIter wird das Ergebnis des rekursiven Aufrufes von fakIter ohne weitere Verarbeitung zurückgegeben. In diesem Fall muss keine Kette verzögerter Aufrufe aufgebaut werden, das Endergebnis entspricht dem Wert der innersten Funktionsauswertung. Diese Optimierung kann vom Compiler durchgeführt werden (tail recursion).

Programm: **(Ausgabe des Programmverlaufs [fakultaetiter_mit_ausgabe.cc])**

```
#include "fcpp.hh"

int faklter( int produkt, int zaehler, int ende )
{
  return cond( zaehler>ende,
               produkt,
               print( "Fak:", zaehler,
                      produkt*zaehler,
                      faklter( produkt*zaehler,
                               zaehler+1,
                               ende ) ) );
}

int fakultaet( int n ) { return faklter( 1, 1, n ); }
int main() { return dump( fakultaet(10) ); }
```

```
Fak: 10 3628800
Fak: 9 362880
Fak: 8 40320
Fak: 7 5040
Fak: 6 720
Fak: 5 120
Fak: 4 24
Fak: 3 6
Fak: 2 2
Fak: 1 1
```

- print (f(x)) = f(x), Wert von f(x) wird gedruckt.
- print (f(x), g(x)) = g(x), Wert von f(x) wird gedruckt.
- print (f(x), g(x), h(x)) = h(x), Wert von f(x) und g(x) wird gedruckt.

Sprechweise: Dies nennt man einen linear iterativen Prozess. Der Zustand des Programmes lässt sich durch eine feste Zahl von Zustandsgrößen beschreiben (hier die Werte von `zaehler` und `produkt`). Es gibt eine Regel wie man von einem Zustand zum nächsten kommt, und es gibt den Endzustand.

Bemerkung:

- Von einem Zustand kann man ohne Kenntnis der Vorgeschichte aus weiterrechnen. Der Zustand fasst alle bis zu diesem Punkt im Programm durchgeführten Berechnungen zusammen.
- Die Zahl der durchlaufenen Zustände ist proportional zu n.
- Die Informationsmenge zur Darstellung des Zustandes ist konstant.
- Bei geeigneter Implementierung ist der Speicherplatzbedarf konstant.
- Beim Lisp-Dialekt Scheme wird diese Optimierung von am Ende aufgerufenen Funktionen (*tail-call position*) im Sprachstandard verlangt.
- Bei anderen Sprachen (auch C++) ist diese Optimierung oft durch Compilereinstellungen erreichbar (nicht automatisch, weil das Debuggen erschwert wird), ist aber nicht Teil des Standards.

- Beide Arten von Prozessen werden durch rekursive Funktionen beschrieben!

Baumrekursion

Beispiel: (Fibonacci-Zahlen)

$$\mathsf{fib}(n) = \begin{cases} 0 & n = 0 \\ 1 & n = 1 \\ \mathsf{fib}(n-1) + \mathsf{fib}(n-2) & n > 1 \end{cases} .$$

Die Folge der Fibonacci Zahlen modelliert (unter anderem) das Wachstum einer Kaninchenpopulation unter vereinfachten Annahmen. Sie ist benannt nach Leonardo di Pisa.[11]

[11] Leonardo di Pisa (auch Fibonacci), etwa 1180–1241, ital. Rechenmeister in Pisa.

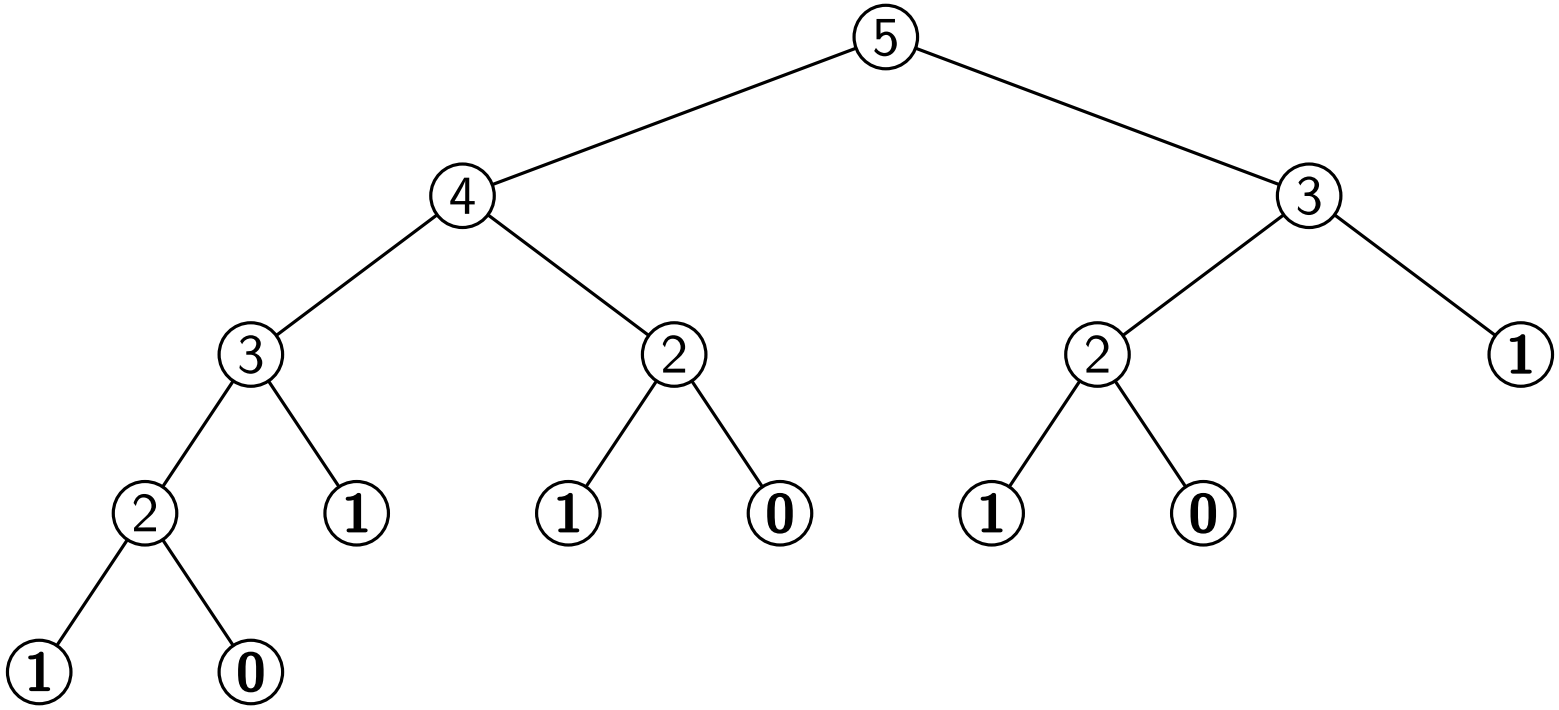
Programm: **(Fibonacci rekursiv [fibonacci.cc])**

```
#include "fcpp.hh"

int fib( int n )
{
  return cond( n==0, 0,
               cond( n==1, 1,
                     fib(n-1) + fib(n-2) ) );
}

int main( int argc, char *argv[] )
{
  return print( fib( readarg_int( argc, argv, 1 ) ) );
}
```

Auswertung von `fib(5)` nach dem Substitutionsmodell:

```
  fib(5)
= +(fib(4),fib(3))
= +(+(fib(3),fib(2)),+(fib(2),fib(1)))
= +(+(+(fib(2),fib(1)),+(fib(1),fib(0))),+(+(fib(1),fib(0)),fib(1)))
= +(+(+(+(fib(1),fib(0)),fib(1)),+(fib(1),fib(0))),+(+(fib(1),fib(0)),fib(1)))
= +(+(+(+(1     ,0     ),1     ),+(1     ,0     )),+(+(1     ,0     ),1     ))
= +(+(+(1               ,1     ),1               ),+(1               ,1     ))
= +(+(2                         ,1               ),2                         )
= +(3                                             ,2                         )
= 5
```

Graphische Darstellung des Aufrufbaumes



`fib(5)` baut auf `fib(4)` und `fib(3)`, `fib(4)` baut auf `fib(3)` und `fib(2)`, usw.

Bezeichnung: Der Rekursionsprozess bei der Fibonaccifunktion heißt daher baumrekursiv.

Frage:

- Wie schnell wächst die Anzahl der Operationen bei der rekursiven Auswertung der Fibonaccifunktion?

- Wie schnell wächst die Fibonaccifunktion selbst?

Antwort: (Wachstum von fib). $F_n := \text{fib}(n)$ erfüllt die lineare 3-Term-Rekursion

$$F_n = F_{n-1} + F_{n-2}$$

Die Lösungen dieser Gleichung sind von der Form $a\lambda_1^n + b\lambda_2^n$, wobei $\lambda_{1/2}$ die Lösungen der quadratischen Gleichung $\lambda^2 = \lambda + 1$ sind, also $\lambda_{1/2} = \frac{1\pm\sqrt{5}}{2}$. Die Konstanten a und b werden durch die Anfangsbedingungen $F_0 = 0, F_1 = 1$ festgelegt und damit ergibt sich

$$F_n = \underbrace{\frac{1}{\sqrt{5}}}_{a}\left(\frac{1+\sqrt{5}}{2}\right)^n \underbrace{-\frac{1}{\sqrt{5}}}_{b}\left(\frac{1-\sqrt{5}}{2}\right)^n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

für große n, da $|\lambda_2| < 1$.

Bemerkung: $\lambda_1 \approx 1.61803$ ist der *goldene Schnitt*.

Antwort: (Aufwand zur rekursiven Berechnung von fib(n))

- Der Gesamtaufwand A_n zur Auswertung von fib (n) ist größer gleich einer Konstante c_1 multipliziert mit der Zahl B_n der Blätter im Berechnungsbaum:

$$A_n \geq c_1 B_n.$$

Die Zahl der Blätter B_n erfüllt die Rekursion:

$$B_0 = 1\,,\ B_1 = 1\,,\ B_n = B_{n-1} + B_{n-2}\,, \quad n > 1$$

woraus man

$$B_n = \mathsf{fib}(n+1) \geq \frac{\lambda_1}{\sqrt{5}}\lambda_1^n - \epsilon_1$$

ersieht (Beachte $B_0 = 1$!). Die Ungleichung gilt für $n \geq N_1(\epsilon_1)$.

- Der Gesamtaufwand A_n zur Auswertung von fib(n) ist kleiner gleich einer Konstante c_2 multipliziert mit der Anzahl G_n der Knoten im Baum:

$$A_n \leq c_2 G_n.$$

Diese erfüllt:

$$G_0 = 1\,,\ G_1 = 1\,,\ G_n = G_{n-1} + G_{n-2} + 1\,, \quad n > 1.$$

Durch die Transformation $G_n = G'_n - 1$ ist dies äquivalent zu

$$G'_0 = 2\,,\ G'_1 = 2\,,\ G'_n = G'_{n-1} + G'_{n-2}\,, \quad n > 1.$$

Mit den Methoden von oben erhält man

$$G'_n = \left(1 + \frac{1}{\sqrt{5}}\right)\lambda_1^n + \left(1 - \frac{1}{\sqrt{5}}\right)\lambda_2^n \leq \left(1 + \frac{1}{\sqrt{5}}\right)\lambda_1^n + \epsilon_2$$

für $n \geq N_2(\epsilon_2)$.

Damit erhalten wir also zusammengefasst:

$$c_1 \frac{\lambda_1}{\sqrt{5}} \lambda_1^n - c_1\epsilon_1 \leq A_n \leq c_2 \left(1 + \frac{1}{\sqrt{5}}\right) \lambda_1^n + c_2\epsilon_2$$

für $n \geq \max(N_1(\epsilon_1), N_2(\epsilon_2))$.

Bemerkung:

- Der Rechenaufwand wächst somit exponentiell.
- Der Speicherbedarf wächst hingegen nur linear in n.

Auch die Fibonaccizahlen kann man iterativ berechnen indem man die aktuelle Summe mitführt:

Programm: (Fibonacci iterativ [fibiter.cc])

```
#include "fcpp.hh"

int fiblter( int letzte, int vorletzte, int zaehler )
{
  return cond( zaehler==0,
               vorletzte,
               fiblter( vorletzte+letzte, letzte, zaehler-1 ) );
}

int fib( int n )
{
  return fiblter( 1, 0, n );
}

int main( int argc, char *argv[] )
{
  return print( fib( readarg_int( argc, argv, 1 ) ) );
}
```

Hier liefert das Substitutionsmodell:

```
  fib(2)
= fibIter(1,0,2)
= cond( 2==0, 0, fibiter(1,1,1))
= fibiter(1,1,1)
= cond( 1==0, 1, fibiter(2,1,0))
= fibIter(2,1,0)
= cond( 0==0, 1, fibiter(3,2,-1))
= 1
```

Bemerkung:

- Man braucht hier offenbar drei Zustandsvariablen.
- Der Rechenaufwand des linear iterativen Prozesses ist proportional zu n, also viel kleiner als der baumrekursive.

Größenordnung

Es gibt eine formale Ausdrucksweise für Komplexitätsaussagen wie „der Aufwand zur Berechnung von `fib(n)` wächst exponentiell“.

Sei n ein Parameter der Berechnung, z. B.

- Anzahl gültiger Stellen bei der Berechnung der Quadratwurzel
- Dimension der Matrix in einem Programm für lineare Algebra
- Größe der Eingabe in Bits

Mit $R(n)$ bezeichnen wir den Bedarf an Resourcen für die Berechnung, z. B.

- Rechenzeit
- Anzahl auszuführender Operationen
- Speicherbedarf

Definition:

- $R(n) = \Omega(f(n))$, falls es von n unabhängige Konstanten c_1, n_1 gibt mit

$$R(n) \geq c_1 f(n) \quad \forall n \geq n_1.$$

- $R(n) = O(f(n))$, falls es von n unabhängige Konstanten c_2, n_2 gibt mit

$$R(n) \leq c_2 f(n) \quad \forall n \geq n_2.$$

- $R(n) = \Theta(f(n))$, falls $R(n) = \Omega(f(n)) \wedge R(n) = O(f(n))$.

Beispiel: $R(n)$ bezeichne den Rechenaufwand der rekursiven Fibonacci-Berechnung:

$$R(n) = \Omega(n)\,, \quad R(n) = O(2^n)\,, \quad R(n) = \Theta(\lambda_1^n)$$

Bezeichnung:

$R(n) = \Theta(1)$	konstante Komplexität
$R(n) = \Theta(\log n)$	logarithmische Komplexität
$R(n) = \Theta(n)$	lineare Komplexität
$R(n) = \Theta(n \log n)$	fast optimale Komplexität
$R(n) = \Theta(n^2)$	quadratische Komplexität
$R(n) = \Theta(n^p)$	polynomiale Komplexität
$R(n) = \Theta(a^n)$	exponentielle Komplexität

Beispiel 1: Telefonbuch

Wir betrachten den Aufwand für das Finden eines Namens in einem Telefonbuch der Seitenzahl n.

Algorithmus: (A1) Blättere das Buch von Anfang bis Ende durch.

Satz: Sei $C_1 > 0$ die (maximale) Zeit, die das Durchsuchen einer Seite benötigt. Der maximale Zeitaufwand $A_1 = A_1(n)$ für Algorithmus A1 ist dann abschätzbar durch

$$A_1(n) = C_1 n$$

Algorithmus: (A2) Rekursives Halbieren.

1. Setze $[a_1 = 1, b_1 = n]$, $i = 1$;

2. Ist $a_i = b_i$ durchsuche Seite a_i; Fertig;

3. Setze $m = (a_i + b_i)/2$ (ganzzahlige Division);

4. Falls Name vor Seite m
   setze $[a_{i+1} = a_i, b_{i+1} = m]$, $i = i + 1$, gehe zu 2.;

5. Falls Name nach Seite m
   setze $[a_{i+1} = m, b_{i+1} = b_i]$, $i = i + 1$, gehe zu 2.;

6. Durchsuche Seite m; Fertig;

Satz: Sei $C_1 > 0$ die (maximale) Zeit, die das Durchsuchen einer Seite benötigt, und $C_2 > 0$ die (maximale) Zeit für die Schritte 3–5. Der maximale Zeitaufwand $A_2 = A_2(n)$ für Algorithmus A2 ist dann abschätzbar durch

$$A_2(n) = C_1 + C_2 \log_2 n$$

Man ist vor allem an der Lösung großer Probleme interessiert. Daher interessiert der Aufwand $A(n)$ für große n.

Satz: Für große Telefonbücher ist Algorithmus 2 „besser“, d. h. der maximale Zeitaufwand ist kleiner.

Beweis:

$$\frac{A_1(n)}{A_2(n)} = \frac{C_1 n}{C_1 + C_2 \log_2 n} = \frac{n}{1 + \frac{C_2}{C_1} \log_2 n} \to +\infty$$

Beobachtung:

- Die genauen Werte von C_1, C_2 sind für diese Aussage unwichtig.
- Für spezielle Eingaben (z. B. Andreas Aalbert) kann auch Algorithmus 1 besser sein.

Bemerkung: Um „Algorithmus 2 ist für große Telefonbücher besser“ zu schließen, reichen die Informationen $A_1(n) = O(n)$ und $A_2(n) = O(\log n)$ aus. Man beachte auch, dass wegen $\log_2 n = \frac{\log n}{\log 2}$ gilt $O(\log_2 n) = O(\log n)$.

Wechselgeld

Aufgabe: Ein gegebener Geldbetrag ist unter Verwendung von Münzen zu 1, 2, 5, 10, 20 und 50 Cent zu wechseln. Wieviele verschiedene Möglichkeiten gibt es dazu?

Beachte: Die Reihenfolge in der wir die Münzen verwenden ist egal.

Idee: Es sei der Betrag a mit n verschiedenen Münzarten zu wechseln. Eine der n Münzarten habe den Nennwert d. Dann gilt:

- Entweder wir verwenden eine Münze mit Wert d, dann bleibt der Rest $a - d$ mit n Münzarten zu wechseln.

- Wir verwenden die Münze mit Wert d $\ddot{u}berhaupt\ nicht$, dann müssen wir den Betrag a mit den verbleibenden $n - 1$ Münzarten wechseln.

Folgerung: Ist $A(a, n)$ die Anzahl der Möglichkeiten den Betrag a mit n Münzarten zu wechseln, und hat Münzart n den Wert d, so gilt

$$A(a, n) = A(a - d, n) + A(a, n - 1)$$

Dies ist ein Beispiel für eine Rekursion in zwei Argumenten.

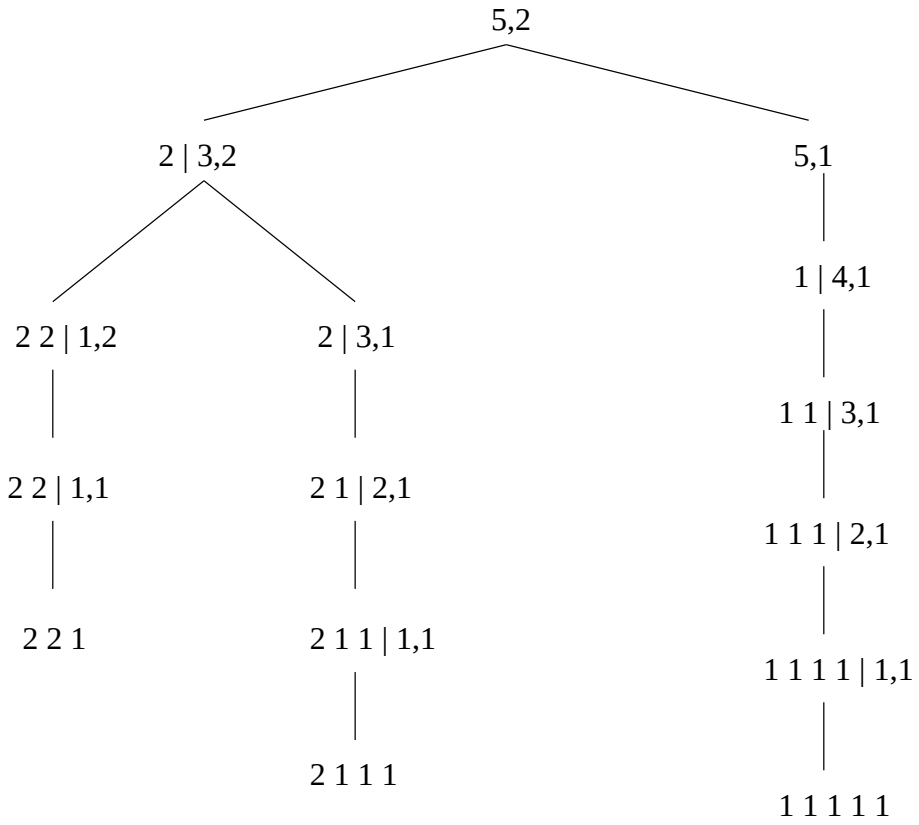
Bemerkung: Es gilt auch:

- $A(0, n) = 1$ für alle $n \geq 0$. Wenn der Betrag a den Wert 0 erreicht hat haben wir den ursprünglichen Betrag gewechselt. ($A(0, 0)$ kann nicht vorkommen).

- $A(a, n) = 0$ falls $a > 0$ und $n = 0$. Der Betrag kann nicht gewechselt werden.

- $A(a, n) = 0$ falls $a < 0$. Der Betrag kann nicht gewechselt werden.

Beispiel: Wechseln von 5 Cent in 1 und 2 Centstücke:



Bemerkung: Dies ist wieder ein baumrekursiver Prozess.

Programm: **(Wechselgeld zählen [wechselgeld.cc])**

```
#include "fcpp.hh"

// uebersetze Muenzart in Muenzwert
int nennwert( int nr )
{
  return
    cond( nr==1, 1,
          cond( nr==2, 2,
                cond( nr==3, 5,
                      cond( nr==4, 10,
                            cond( nr==5, 20,
                                  cond( nr==6, 50, 0 )
                                ) ) ) ) );
}

int wg( int betrag, int muenzarten )
{
  return cond( betrag==0, 1,
```

```
                cond( betrag<0 || muenzarten==0, 0,
                      wg( betrag, muenzarten-1 ) +
                      wg( betrag - nennwert(muenzarten),
                          muenzarten ) ) );
}

int wechselgeld( int betrag )
{
  return wg( betrag, 6 );
}

int main( int argc, char *argv[] ) {
  return print( wechselgeld( readarg_int( argc, argv, 1 ) ) );
}
```

Hier einige Resultate:

```
wechselgeld(50)  = 451
wechselgeld(100) = 4562
wechselgeld(200) = 69118
wechselgeld(300) = 393119
```

Bemerkung: Ein iterativer Lösungsweg ist hier nicht ganz so einfach.

Der größte gemeinsame Teiler

Definition: Als den größten gemeinsamen Teiler (ggT) zweier Zahlen $a, b \in \mathbb{N}_0$ bezeichnen wir die größte natürliche Zahl, die sowohl a als auch b ohne Rest teilt.

Bemerkung: Den ggT braucht man etwa um rationale Zahlen zu kürzen:

$$\frac{91}{287} = \frac{13}{41}, \ \text{ggT}(91, 287) = 7.$$

Idee: Zerlege beide Zahlen in Primfaktoren, der ggT ist dann das Produkt aller gemeinsamer Faktoren. Leider: sehr aufwendig.

Effizienter: Euklidscher[12] Algorithmus. Dieser basiert auf folgenden Überlegungen:

Bezeichnung: Seien $a, b \in \mathbb{N}$ (ohne Null). Dann gibt es Zahlen q, r so dass

[12] Euklid von Alexandria, ca. 360–280 v. Chr., bedeutender griechischer Mathematiker.

$a = q \cdot b + r$ mit $q \in \mathbb{N}_0$ und $0 \leq r < b$. Wir schreiben $a \bmod b$ für den Rest r. Wenn $r = 0$, so schreiben wir $b|a$.

Bemerkung:

1. Wir verlangen $a + b > 0$.

2. Falls $b = 0$ und $a > 0$, so ist $\mathsf{ggT}(a, b) = a$. (Jedes $n \in \mathbb{N}$ teilt 0).

3. Für jeden Teiler s von a und b gilt $\frac{a}{s} = q\frac{b}{s} + \frac{r}{s} \in \mathbb{N}$. Wegen $\frac{a}{s} \in \mathbb{N}$ und $\frac{b}{s} \in \mathbb{N}$ nach Voraussetzung muss auch $\frac{r}{s} \in \mathbb{N}$ gelten, d. h. s ist auch Teiler von r!

   Somit gilt $\mathsf{ggT}(a, b) = \mathsf{ggT}(b, r)$.

Somit haben wir folgende Rekursion bewiesen:

$$\mathsf{ggT}(a, b) = \begin{cases} a & \text{falls } b = 0 \\ \mathsf{ggT}(b, a \bmod b) & \text{sonst} \end{cases}$$

Programm: **(Größter gemeinsamer Teiler [ggT.cc])**

```
#include "fcpp.hh"

int ggT( int a, int b )
{
  return cond( b==0, a, ggT( b, a % b ) );
}

int main( int argc, char *argv[] )
{
  return print( ggT( readarg_int( argc, argv, 1 ),
                     readarg_int( argc, argv, 2 ) ) );
}
```

Hier die Berechnung von `ggt(91,287)`

```
  ggT(91,287)     #  91 = 0*287 + 91
= ggT(287,91)     # 287 = 3*91 + 14
= ggT(91,14)      #  91 = 6*14 + 7
= ggT(14,7)       #  14 = 2*7 + 0
= ggT(7,0)
= 7
```

- Terminiert das Verfahren immer?
- Wie schnell terminiert es?

Bemerkung:

- Im ersten Schritt ist $91 = 0 \cdot 287 + 91$, also werden die Argumente gerade vertauscht.
- Der Berechnungsprozess ist iterativ, da nur ein fester Satz von Zuständen mitgeführt werden muss.

Satz: Der Aufwand von `ggT(a,b)` ist $O(\log n)$, wobei $n = \min(a, b)$.

Beweis: Ausgehend von der Eingabe $a_0 = a, b_0 = b$, $a, b \in \mathbb{N}_0$, $a + b > 0$, erzeugt der Euklidsche Algorithmus eine Folge von Paaren

$$(a_i, b_i), \qquad i \in \mathbb{N}_0.$$

Dabei gilt nach Konstruktion

$$a_{i+1} = b_i, \qquad b_{i+1} = a_i \bmod b_i.$$

Wir beweisen nun einige Eigenschaften dieser Folge.

1. Es gilt $b_i < a_i$ für alle $i \geq 1$. Wir zeigen dies in zwei Schritten.

   α. Sei bereits $b_i < a_i$, dann gilt

   $$a_i = q_i b_i + r_i \quad \text{mit } 0 \leq r_i < b_i.$$

   Da $a_{i+1} = b_i$ und $b_{i+1} = r_i$ gilt offensichtlich

   $$b_{i+1} = r_i < b_i = a_{i+1}.$$

   β. Ist $b_0 < a_0$ dann gilt wegen α. auch $b_1 < a_1$. Bleiben also die Fälle $b_0 = a_0$ und $b_0 > a_0$:

   $$\begin{aligned} b_0 = a_0 \quad &\Rightarrow a_0 = 1 \cdot b_0 + 0 \Rightarrow b_1 = 0 < b_0 = a_1. \\ b_0 > a_0 \quad &\Rightarrow a_0 = 0 \cdot b_0 + a_0 \Rightarrow b_1 = a_0 < b_0 = a_1. \end{aligned}$$

2. Nun können wir bereits zeigen, dass der Algorithmus terminieren muss. Wegen 1. gilt

$$b_{i+1} < a_{i+1} = b_i < a_i, \quad \text{für } i \geq 1,$$

mithin ist also die Folge der b_i (und auch der a_i) streng monoton fallend. Wegen $b_i \in \mathbb{N}_0$ impliziert $b_{i+1} < b_i$ dass $b_{i+1} \leq b_i - 1$.

Andererseits ist $b_i \geq 0$ für alle $i \geq 0$ da $b_0 \geq 0$ und $b_{i+1} = a_i \bmod b_i$. Somit muss irgendwann $b_i = 0$ gelten und der Algorithmus terminiert.

3. Sei $b_i < a_i$. Dann gilt $b_{i+2} < a_{i+2} < a_i/2$. Dies ist also eine Behauptung über die Konvergenzgeschwindigkeit. Wir unterscheiden zwei Fälle.

   I. Sei $b_i \leq a_i/2$. Dann gilt $a_i = q_i \cdot b_i + r_i$ mit $0 \leq r_i < b_i \leq a_i/2$, also $b_{i+1} = r_i < b_i = a_{i+1} \leq a_i/2$.
   Im nächsten Schritt gilt dann $a_{i+1} = q_{i+1} \cdot b_{i+1} + r_{i+1}$ mit

$$b_{i+2} = r_{i+1} < b_{i+1} = a_{i+2} < b_i \leq a_i/2.$$

   Somit gilt $b_{i+2} < a_{i+2} < a_i/2$.

II. Sei $b_i > a_i/2$. Dann gilt $a_i = 1 \cdot b_i + (a_i - b_i)$, also $q_i = 1$, $r_i = a_i - b_i$. Damit gilt $b_{i+1} = r_i = a_i - b_i < a_i/2$ und $a_{i+1} = b_i > a_i/2$ (nach Vor.). Im nächsten Schritt gilt nun wieder $a_{i+1} = q_{i+1} \cdot b_{i+1} + r_{i+1}$ mit

$$b_{i+2} = r_{i+1} < b_{i+1} = a_{i+2} < a_i/2,$$

also ebenfalls $b_{i+2} < a_{i+2} < a_i/2$.

Damit ist gezeigt, dass a_i und b_i nach zwei Schritten noch höchstens halb so groß sind. Da $a_i, b_i \in \mathbb{N}_0$ sind höchstens $2\log_2(\min(a_0, b_0))$ Halbierungen möglich bis b_i den Wert 0 erreicht.

Zahlendarstellung im Rechner

In der Mathematik gibt es verschiedene Zahlenmengen:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Diese Mengen enthalten alle unendlich viele Elemente, im Computer entsprechen die diversen Datentypen jedoch nur endlichen Mengen.

Um Zahlen aus $\mathbb{N}$ darzustellen, benutzt man ein Stellenwertsystem zu einer Basis $\beta \geq 2$ und Ziffern $a_i \in \{0, \ldots, \beta - 1\}$

Dann bedeutet

$$(a_{n-1}a_{n-2}\ldots a_1 a_0)_\beta \equiv \sum_{i=0}^{n-1} a_i \beta^i$$

Dabei ist n die Wortlänge. Es sind somit die folgenden Zahlen aus $\mathbb{N}_0$ darstellbar:

$$0, 1, \ldots, \beta^n - 1$$

Am häufigsten wird $\beta = 2$, das Binärsystem, verwendet.

Umgang mit negativen Zahlen

In der Mathematik ist das Vorzeichen eine separate Information welche 1 Bit zur Darstellung benötigt.

Im Rechner wird bei ganzen Zahlen zur Basis $\beta = 2$ eine andere Darstellung gewählt, die Zweierkomplementdarstellung.

Früher war auch das Einerkomplement gebräuchlich.

Einer- und Zweierkomplement

Definition: Sei $(a_{n-1}a_{n-2}\ldots a_1a_0)_2$ die Binärdarstellung von $a \in [0, 2^n - 1]$. Dann heisst

$$e_n(a) = e_n((a_{n-1}a_{n-2}\ldots a_1a_0)_2) = (\bar{a}_{n-1}\bar{a}_{n-2}\ldots \bar{a}_1\bar{a}_0)_2$$

das Einerkomplement von a, wobei $\bar{a}_i = 1 - a_i$.

Definition: Sei $a \in [0, 2^n - 1]$. Dann heisst

$$z_n(a) = 2^n - a$$

das Zweierkomplement von a.

Es gilt: $e_n(e_n(a)) = a$ und $z_n(z_n(a)) = a$!

Das Zweierkomplement einer Zahl kann sehr einfach und effizient berechnet werden.

Sei $a \in [0, 2^n - 1]$. Dann folgt aus der Identität

$$a + e_n(a) = 2^n - 1$$

die Formel

$$z_n(a) = 2^n - a = e_n(a) + 1.$$

Somit wird keine Subtraktion benötigt sondern es genügt das Einerkomplement und eine Addition von 1. (Und das kann noch weiter vereinfacht werden).

Definition: Die Zweierkomplementdarstellung einer Zahl ist eine bijektive Abbildung

$$d_n : [-2^{n-1}, 2^{n-1} - 1] \to [0, 2^n - 1]$$

welche definiert ist als

$$d_n(a) = \begin{cases} a & 0 \leq a < 2^{n-1} \\ 2^n - |a| & -2^{n-1} \leq a < 0 \end{cases} .$$

Die negativen Zahlen $[-2^{n-1}, -1]$ werden damit auf den Bereich $[2^{n-1}, 2^n - 1]$ positiver Zahlen abgebildet.

Sei $d_n(a) = (x_{n-1}x_{n-2}\ldots x_1x_0)_2$ dann ist a positiv falls $x_{n-1} = 0$ und a negativ falls $x_{n-1} = 1$.

Es gilt $d_n(-1) = 2^n - 1 = (1, \ldots, 1)_2$ und $d_n(-2^{n-1}) = 2^n - 2^{n-1} = 2^{n-1}(2-1) = 2^{n-1} = (1, 0, \ldots, 0)_2$.

Operationen mit der Zweierkomplementdarstellung

Die Ein-/Ausgabe von ganzen Zahlen erfolgt (1) im Zehnersystem und (2) mittels separatem Vorzeichen.

Bei der Eingabe einer ganzen Zahl wird diese in das Zweierkomplement umgewandelt:

- Lese Betrag und Vorzeichen ein und teste auf erlaubten Bereich
- Wandle Betrag in das Zweiersystem
- Für negative Zahlen berechne das Zweierkomplement

Bei der Ausgabe gehe umgekehrt vor.

Im folgenden benötigen wir noch eine weitere Operation.

Definition: Sei $x = (x_{m-1}x_{m-2}\dots x_1x_0)_2$ eine m-stellige Binärzahl. Dann ist

$$s_n((x_{m-1}x_{m-2}\dots x_1x_0)_2) = \begin{cases} (x_{m-1}x_{m-2}\dots x_1x_0)_2 & m \leq n \\ (x_{n-1}x_{n-2}\dots x_1x_0)_2 & m > n \end{cases}$$

die Beschneidung auf n-stellige Binärzahlen.

Die Addition von Zahlen in der Zweierkomplementdarstellung gelingt mit

Satz: Sei $n \in \mathbb{N}$ und $a, b, a+b \in [-2^{n-1}, 2^{n-1}-1]$. Dann gilt

$$d_n(a+b) = s_n(d_n(a) + d_n(b)).$$

Es genügt eine einfache Addition. Beachtung der Vorzeichen entfällt!

Beweis. Wir unterscheiden drei Fälle (mittlerer Fall steht für zwei).

$\underline{a, b \geq 0.}$ Damit ist auch $a + b \geq 0$. Also

$$s_n(d_n(a) + d_n(b)) = s_n(a + b) = a + b = d_n(a + b).$$

$\underline{a < 0, b \geq 0.}$ $a + b$ kann positiv oder negativ sein. Mit $a = -|a|$:

$$\begin{aligned} s_n(d_n(a) + d_n(b)) &= s_n(2^n - |a| + b) = s_n(2^n + (a + b)) \\ &= \begin{cases} a + b & 0 \leq a + b < 2^{n-1} \\ 2^n - |a + b| & -2^{n-1} \leq a + b < 0 \end{cases}. \end{aligned}$$

$\underline{a, b < 0.}$ Damit ist auch $a + b < 0$ und $2^n - |a + b| < 2^n$:

$$\begin{aligned} s_n(d_n(a) + d_n(b)) &= s_n(2^n - |a| + 2^n - |b|) = s_n(2^n + 2^n + a + b) \\ &= s_n(2^n + 2^n - |a + b|) = 2^n - |a + b|. \end{aligned}$$

Beispiel: (Zweierkomplement) Für $n = 3$ setze

0	=	000	-1	=	111
1	=	001	-2	=	110
2	=	010	-3	=	101
3	=	011	-4	=	100

Solange der Zahlenbereich nicht verlassen wird, klappt die normale Arithmetik ohne Beachtung des Vorzeichens:

3	$\rightarrow$	011
-1	$\rightarrow$	111
2	$\rightarrow$	[1]010

Die Negation einer Zahl in Zweierkomplementdarstellung.

Satz: Sei $n \in \mathbb{N}$ und $a \in [-2^{n-1}+1, 2^{n-1}-1]$. Dann gilt

$$d_n(-a) = s_n(2^n - d_n(a)) = s_n(e_n(d_n(a)) + 1).$$

Beweis.

$\underline{a = 0.}$ $d_n(-0) = d_n(0) = s_n(2^n - d_n(0))$. Nur dieser Fall braucht die Anwendung von s_n.

$\underline{0 < a < 2^{n-1}.}$ Also ist $-a < 0$ und somit

$$d_n(-a) = 2^n - |a| = 2^n - a = 2^n - d_n(a) = s_n(2^n - d_n(a)).$$

$\underline{-2^{n-1} < a < 0.}$ Also ist $-a > 0$ und somit

$$d_n(-a) = d_n(|a|) = |a| = 2^n - (2^n - |a|) = 2^n - d_n(a) = s_n(2^n - d_n(a)).$$

Schließlich behandeln wir noch die Subtraktion.

Diese wird auf die Addition zurückgeführt:

$$
\begin{aligned}
d_n(a-b) &= d_n(a+(-b)) && a-b = a+(-b)\\
&= s_n(d_n(a)+d_n(-b)) && \text{Satz über Addition}\\
&= s_n(d_n(a)+s_n(2^n-d_n(b))) && \text{Satz über Negation.}
\end{aligned}
$$

Natürlich vorausgesetzt, dass alles im erlaubten Bereich ist.

Gebräuchliche Zahlenbereiche in C++

$\beta = 2$ und $n = 8, 16, 32$:

`char`	-128. . . 127
`unsigned char`	0. . . 255
`short`	-32768. . . 32767
`unsigned short`	0. . . 65535
`int`	-2147483648. . . 2147483647
`unsigned int`	0. . . 4294967295

Bemerkung: Die genaue Größe legt der C++ Standard nicht fest, diese kann aber abgefragt werden.

Bemerkung: Achtung: bei Berechnungen mit ganzen Zahlen führt Überlauf, d. h. das Verlassen des darstellbaren Zahlenbereichs üblicherweise nicht zu Fehlermeldungen. Es liegt in der Verantwortung des Programmierers solche Fehler zu vermeiden.

Darstellung reeller Zahlen

Neben den Zahlen aus $\mathbb{N}$ und $\mathbb{Z}$ sind in vielen Anwendungen auch reelle Zahlen $\mathbb{R}$ von Interesse. Wie werden diese im Computer realisiert?

Festkommazahlen

Eine erste Idee ist die Festkommazahl. Hier interpretiert man eine gewisse Zahl von Stellen als $nach\ dem\ Komma$, d. h.

$$(a_{n-1}a_{n-2}\dots a_q.a_{q-1}\dots a_0)_\beta \equiv \sum_{i=0}^{n-1} a_i\beta^{i-q}$$

Beispiel: Bei $\beta = 2, q = 3$ hat man drei Nachkommastellen, kann also in Schritten von $1/8$ auflösen.

Bemerkung:

- Jede Festkommazahl ist rational, somit können irrationale Zahlen nicht exakt dargestellt werden.

- Selbst einfache rationale Zahlen können je nach Basis nicht exakt dargestellt werden. So kann $0.1 = 1/10$ mit einer Festkommazahl zur Basis $\beta = 2$ für kein n exakt dargestellt werden.

- Das Ergebnis elementarer Rechenoperationen $+, -, *, /$ muss nicht mehr darstellbar sein.

- Festkommazahlen werden nur in Spezialfällen verwendet, etwa um mit Geldbeträgen zu rechnen. In vielen anderen Fällen ist die im nächsten Abschnitt dargestellte Fließkommaarithmetik brauchbarer.

Fließkommaarithmetik

Vor allem in den Naturwissenschaften wird die Fließkommaarithmetik (Gleitkommaarithmetik) angewendet. Eine Zahl wird dabei repräsentiert als

$$\pm \left(a_0 + a_1 \beta^{-1} + \ldots + a_{n-1} \beta^{-(n-1)} \right) \times \beta^e$$

Die Ziffern a_i bilden die Mantisse und e ist der Exponent (eine ganze Zahl gegebener Länge). Wieder wird $\beta = 2$ am häufigsten verwendet. Das Vorzeichen ist ein zusätzliches Bit.

Typische Wortlängen

`float`: 23 Bit Mantisse, 8 Bit Exponent, 1 Bit Vorzeichen entsprechen

$$23 \cdot \log_{10} 2 = 23 \cdot \frac{\log 2}{\log 10} \approx 23 \cdot 0.3 \approx 7$$

dezimalen Nachkommastellen in der Mantisse.

`double`: 52 Bit Mantisse, 11 Bit Exponent, 1 Bit Vorzeichen entsprechen $52 \cdot 0.3 \approx 16$ dezimalen Nachkommastellen in der Mantisse.

Referenz: Genaueres findet man im IEEE-Standard 754 (floating point numbers).

Fehler in der Fließkommaarithmetik

Darstellungsfehler $\beta = 10$, $n = 3$: Die reelle Zahl 3.14159 wird auf 3.14×10^0 gerundet. Der Fehler beträgt maximal 0.005, man sagt $0.5ulp$, ulp heißt $units\ last\ place$.

Bemerkung:

- Wenn solche fehlerbehafteten Daten als Anfangswerte für Berechnungen verwendet werden, können die Anfangsfehler erheblich vergrößert werden.
- Durch Rundung können weitere Fehler hinzukommen.
- Vor allem bei der Subtraktion kann es zum Problem der Auslöschung kommen, wenn beinahe gleichgroße Zahlen voneinander abgezogen werden.

Beispiel: Berechne $b^2 - 4ac$ in $\beta = 10$, $n = 3$ für $b = 3.34, a = 1.22, c = 2.28$. Eine exakte Rechnung liefert

$$3.34 \cdot 3.34 - 4 \cdot 1.22 \cdot 2.28 = 11.1556 - 11.1264 = 0.0292$$

Mit Rundung der Zwischenergebnisse ergibt sich dagegen

$$\ldots 11.2 - 11.1 = 0.1$$

Der absolute Fehler ist somit $0.1 - 0.0292 = 0.0708$. Damit ist der relative Fehler $\frac{0.0708}{0.0292} = 240\%$! Nicht einmal eine Stelle des Ergebnisses $1.00 \cdot 10^{-1}$ ist korrekt!

Typkonversion

Im Ausdruck `5/3` ist „/“ die ganzzahlige Division ohne Rest. Bei `5.0/3.0` oder `5/3.0` oder `5.0/3` wird hingegen eine Fließkommadivision durchgeführt.

Will man eine gewisse Operation erzwingen, kann man eine explizite Typkonversion einbauen:

`((double) x)/3`	Fließkommadivision
`((int) y)/((int) 3)`	Ganzzahldivision

Wurzelberechnung mit dem Newtonverfahren

Problem: $f : \mathbb{R} \to \mathbb{R}$ sei eine „glatte“ Funktion, $a \in \mathbb{R}$. Wir wollen die Gleichung

$$f(x) = a$$

lösen.

Beispiel: $f(x) = x^2 \rightsquigarrow$ Berechnung von Quadratwurzeln.

Mathematik: $\sqrt{a}$ ist die positive Lösung von $x^2 = a$.

Informatik: Will Algorithmus zur Berechnung des Zahlenwerts von $\sqrt{a}$.

Ziel: Konstruiere ein Iterationsverfahren mit folgender Eigenschaft: zu einem Startwert $x_0 \approx x$ finde $x_1, x_2, \ldots$, welche die Lösung x immer besser approximieren.

Definition: (Taylorreihe)

$$f(x_n + h) = f(x_n) + hf'(x_n) + \frac{h^2}{2} f''(x_n) + \dots$$

Wir vernachlässigen nun den $O(h^2)$-Term ($|f''(x)| \leq C$, kleines h) und verlangen $f(x_n + h) \approx f(x_n) + hf'(x_n) \stackrel{!}{=} a$. Dies führt zu

$$h = \frac{a - f(x_n)}{f'(x_n)}$$

und somit zur Iterationsvorschrift

$$x_{n+1} = x_n + \frac{a - f(x_n)}{f'(x_n)}.$$

Beispiel: Für die Quadratwurzel erhalten wir mit $f(x) = x^2$ und $f'(x) = 2x$ die Vorschrift

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right).$$

Abbruchkriterium: $|f(x_n) - a| < \varepsilon$ für eine vorgegebene (kleine) Zahl ε.

Programm: **(Quadratwurzelberechnung [newton.cc])**

```
#include "fcpp.hh"

bool gut_genug( double xn, double a ) {
  return fabs( xn*xn - a ) <= 1e-15;
}

double wurzellter( double xn, double a ) {
  return cond( gut_genug( xn, a ),
               xn,
               wurzellter( 0.5*(xn + a/xn), a ) );
}

double wurzel( double a )
{
  return wurzellter( 1.0, a );
}
```

```
int main( int argc, char *argv[] )
{
  return
    print( wurzel( readarg_double( argc, argv, 1 ) ) );
}
```

Hier ist die Auswertung der Wurzelfunktion im Substitutionsmodell (nur die Aufrufe von `wurzelIter` sind dargestellt):

```
  wurzel(2)
= wurzelIter(1,2)
= wurzelIter(1.5,2)
= wurzelIter(1.4166666666666667,2)
= wurzelIter(1.4142156862745098,2)
= wurzelIter(1.4142135623746898,2)
= wurzelIter(1.4142135623730951,2)
= 1.4142135623730951
```

Bemerkung:

- Die print-Funktion sorgt dafür, dass 16 Stellen bei Fließkommazahlen ausgegeben werden.

- Unter gewissen Voraussetzungen an f kann man zeigen, dass sich die Zahl der gültigen Ziffern mit jedem Schritt verdoppelt.

Fortgeschrittenere funktionale Programmierung

Funktionen in der Mathematik

Definition: Eine Funktion $f : X \to Y$ ordnet jedem Element einer Menge X genau ein Element der Menge Y zu.

In der Mathematik ist es nun durchaus üblich, nicht nur einfache Beispiele wie etwa $f : X \to Y$ mit $X = Y = \mathbb{R}$ zu betrachten. Im Gegenteil: in wichtigen Fällen sind X und/oder Y selbst Mengen von Funktionen.

Beispiele:

- Ableitung: Funktionen $\to$ Funktionen; $X = C^1([a, b])$, $Y = C^0([a, b])$
- Stammfunktion: Funktionen $\to$ Funktionen
- Integraler Mittelwert: Funktionen $\to$ Zahlen

In C und C++ ist es ebenfalls möglich Funktionen als Argumente an Funktionen zu übergeben. Dazu stehen zwei Möglichkeiten zur Verfügung:

- Funktionszeiger (behandeln wir nicht)
- Funktoren (behandeln wir im Rahmen der Objektorientierung)

Bei kompilierten Sprachen werden alle Funktionen zur Übersetzungszeit erzeugt (aber arbeiten auf zur Laufzeit erzeugten Daten).

Interpretierte Sprachen erlauben auch die Erzeugung des Codes selbst zur Laufzeit.

Beispiel: Beispiel mit sog. Lambdaausdruck aus C++ 11:

```
#include "fcpp.hh"

typedef int (*F)( int n ); // Datentyp Funktionszeiger

int apply( F f, int arg )  // apply wendet f auf arg an
{
  return f(arg);
}

int main( int argc, char *argv[] )
{
  return print( apply( [](  int n ){ return n+5; }, // anonyme Fkt.
                       readarg_int( argc, argv, 1 ) ) );
}
```

Warum funktionale Programmierung?

Mathematisch am besten verstanden.

Vorteilhaft, wenn Korrektheit von Programmen gezeigt werden soll.

Die funktionale Programmierung kommt mit wenigen Konzepten aus C++ aus: Auswertung von Ausdrücken, Funktionsdefinition, cond-Funktion.

Bestimmte Probleme lassen sich mit rekursiv formulierten Algorithmen (und nur mit diesen) sehr elegant lösen.

Funktionales Programmieren ist nicht für alle Situationen die beste Wahl! Zum Beispiel legt die Interaktion mit der Außenwelt oder ihre effiziente Nachbildung oft andere Paradigmen nahe.

Ungeschickte rekursive Formulierungen können zu sehr langen Laufzeiten führen. Geeignete Formulierung und Endrekursion können dies vermeiden.

Prozedurale Programmierung

Bisher besteht unser Berechnungsmodell aus folgenden Bestandteilen:

1. Auswertung von zusammengesetzten Ausdrücken aus Zahlen, Funktionsaufrufen und Selektionen.

2. Konstruktion neuer Funktionen.

Namen treten dabei in genau zwei Zusammenhängen auf:

1. Als Symbole für Funktionen.

2. Als formale Parameter in Funktionen.

Im Substitutionsmodell werden bei der Auswertung einer Funktion die formalen Parameter im Rumpf durch die aktuellen Werte ersetzt und dann der Rumpf ausgewertet.

Unter Vernachlässigung von Resourcenbeschränkungen (endlich große Zahlen, endlich viele rekursive Funktionsauswertungen) kann man zeigen, dass dieses Berechnungsmodell äquivalent zur Turingmaschine ist.

In der Praxis erweist es sich als nützlich, weitere Konstruktionselemente einzuführen um einfachere, übersichtlichere und wartbarere Programme schreiben zu können.

Der Preis dafür ist, dass wir uns von unserem einfachen Substitutionsmodell verabschieden müssen!

Lokale Variablen und die Zuweisung

Konstanten

In C++ kann man konstante Werte wie folgt mit Namen versehen:

Beispiel:

```
float umfang( float r )
{
  const double pi = 3.14159265;
  return 2*r*pi;
}

int hochacht( int x )
{
  const int x2 = x*x;          // jetzt gibt es ein x2
  const int x4 = x2*x2;        // nun ein x4
  return x4*x4;
}
```

Bemerkung:

- Wir können uns vorstellen, dass einem Namen ein Wert zugeordnet wird.
- Einem formalen Parameter wird der Wert bei Auswertung der Funktion zugeordnet.
- Einer Konstanten kann nur einmal ein Wert zugeordnet werden.
- Die Auswertung solcher Funktionsrümpfe erfordert eine Erweiterung des Substitutionsmodells:
  - Ersetze formale Parameter durch aktuelle Parameter.
  - Erzeuge (der Reihe nach !) die durch die Zuweisungen gegebenen Name-Wert Zuordnungen und ersetze neue Namen im Rest des Rumpfes durch den Wert.

Variablen

C++ erlaubt aber auch, die Zuordnung von Werten zu Namen zu ändern:

Beispiel: (Variablen)

```
int hochacht( int x )
{
  int y = x*x;   // Zeile 1: Definition/Initialisierung
  y = y*y;       // Zeile 2: Zuweisung
  return y*y;
}
```

Bemerkung:

- Zeile 1 im Funktionsrumpf definiert eine Variable `y`, die Werte vom Typ `int` annehmen kann und initialisiert diese mit dem Wert des Ausdrucks `x*x`.
- Zeile 2 nennt man eine Zuweisung. Die links des = stehende Variable erhält den Wert des rechts stehenden Ausdrucks als neuen Wert.
- Beachte, dass der boolsche Operator „ist gleich“ (also die Abfrage nach Gleichheit) in C++ durch == notiert wird!
- Der Wert von `y` wird erst geändert, nachdem der Ausdruck rechts ausgewertet wurde.
- Der Typ einer Variablen kann aber nicht geändert werden!

Problematik der Zuweisung

Beispiel:

```
int bla( int x )
{
  int y = 3;                  // Zeile 1
  const int x1 = y*x;         // Zeile 2
  y = 5;                      // Zeile 3
  const int x2 = y*x;         // Zeile 4
  return x1*x2;               // Zeile 5
}
```

Bemerkung:

- Obwohl `x1` und `x2` durch denselben Ausdruck `y*x` definiert werden, haben sie im allgemeinen verschiedene Werte.

- Dies bedeutet das Versagen des Substitutionsmodells, bei dem ein Name im ganzen Funktionsrumpf durch seinen Wert ersetzt werden kann.

- Die Namensdefinitionen und Zuweisungen werden der Reihe nach abgearbeitet. Das Ergebnis hängt auch von dieser Reihenfolge ab. Dagegen war die Reihenfolge der Auswertung von Ausdrücken im Substitutionsmodell egal.

Umgebungsmodell

Wir können uns die Belegung der Variablen als Abbildung bzw. Tabelle vorstellen, die jedem Namen einen Wert zuordnet:

$$w : \{ \text{ Menge der gültigen Namen } \} \to \{ \text{ Menge der Werte } \}.$$

Beispiel: Abbildung w bei Aufruf von `bla(4)` nach Zeile 4

Name	Typ	Wert
`x`	`int`	4
`y`	`int`	5
`x1`	`int`	12
`x2`	`int`	20

Definition: Der Ort, an dem diese Abbildung im System gespeichert wird, heißt Umgebung. Die Abbildung w heißt auch Bindungstabelle. Man sagt, w bindet einen Namen an einen Wert.

Bemerkung:

- Ein Ausdruck wird in Zukunft immer relativ zu einer Umgebung ausgewertet, d. h. nur Ausdruck und Umgebung zusammen erlauben die Berechnung des Wertes eines Ausdruckes.

- Die Zuweisung können wir nun als Modifikation der Bindungstabelle verstehen: nach der Ausführung von y=5 gilt $w(\mathrm{y}) = 5$.

- Die Bindungstabelle ändert sich dynamisch während der Programmausführung. Um herauszufinden „was ein Programm tut“ muss sich der Programmierer die fortwährende Entwicklung der Bindungstabelle vorstellen.

Syntax von Variablendefinition und Zuweisung

Syntax:

$$
\begin{array}{lll}
<\text{Def}> & ::= & <\text{ConstDef}> \mid <\text{VarDef}> \\
<\text{ConstDef}> & ::= & \underline{\text{const}} <\text{Typ}> <\text{Name}> \equiv <\text{Ausdruck}> \\
<\text{VarDef}> & ::= & <\text{Typ}> <\text{Name}> [\equiv <\text{Ausdruck}>]
\end{array}
$$

Syntax:

$$
<\text{Zuweisung}> \quad ::= \quad <\text{Name}> \equiv <\text{Ausdruck}>
$$

Bemerkung:

- Wir erlauben zunächst Variablendefinitionen nur innerhalb von Funktionsdefinitionen. Diese Variablen bezeichnet man als lokale Variablen.

- Bei der Definition von Variablen $kann$ die Initialisierung weggelassen werden. In diesem Fall ist der Wert der Variablen bis zur ersten Zuweisung unbestimmt. Aber: Fast immer ist es empfehlenswert, auch Variablen gleich bei der Definition zu initialisieren!

Lokale Umgebung

Wie sieht die Umgebung im Kontext mehrerer Funktionen aus?

Programm:

```
int g( int x )
{
  int y = x*x;
  y = y*y;
  return h( y*(x+y) );
}

int h( int x )
{
  return cond( x<1000, g(x), 88 );
}
```

Es gilt folgendes:

- Jede Auswertung einer Funktion erzeugt eine eigene lokale Umgebung. Mit Beendigung der Funktion wird diese Umgebung wieder vernichtet!
- Zu jedem Zeitpunkt der Berechnung gibt es eine aktuelle Umgebung. Diese enthält die Bindungen der Variablen der Funktion, die gerade ausgewertet wird.
- In Funktion `h` gibt es keine Bindung für `y`, auch wenn `h` von `g` aufgerufen wurde.
- Wird eine Funktion n mal rekursiv aufgerufen, so existieren n verschiedene Umgebungen für diese Funktion.

Bemerkung: Man beachte auch, dass eine Funktion kein Gedächtnis hat: wird sie mehrmals mit gleichen Argumenten aufgerufen, so sind auch die Ergebnisse gleich. Diese fundamentale Eigenschaft funktionaler Programmierung ist also (bisher) noch erhalten.

Bemerkung: Tatsächlich wäre obiges Konstrukt auch nach Einführung einer `main`-Funktion nicht kompilierbar, weil die Funktion `h` beim Übersetzen von `g` noch nicht bekannt ist. Um dieses Problem zu umgehen, erlaubt C++ die vorherige Deklaration von Funktionen. In obigem Beispiel könnte dies geschehen durch Einfügen der Zeile

```
int h( int x );
```

vor die Funktion `g`.

Anweisungsfolgen (Sequenz)

- Funktionale Programmierung bestand in der Auswertung von Ausdrücken.
- Jede Funktion hatte nur eine einzige Anweisung (`return`).
- Mit Einführung von Zuweisung (oder allgemeiner Nebeneffekten) macht es Sinn, die *Ausführung mehrer Anweisungen* innerhalb von Funktionen zu erlauben. Diesen Programmierstil nennt man auch *imperative Programmierung*.

Erinnerung: Wir kennen schon eine Reihe wichtiger Anweisungen:

- Variablendefinition (ist in C++ eine Anweisung, nicht aber in C),
- Zuweisung,
- `return`-Anweisung in Funktionen.

Bemerkung:

- Jede Anweisung endet mit einem Semikolon.
- Überall wo eine Anweisung stehen darf, kann auch eine durch geschweifte Klammern eingerahmte Folge ($Sequenz$) von Anweisungen stehen.
- Auch die leere Anweisung ist möglich indem man einfach ein Semikolon einfügt.
- Anweisungen werden der Reihe nach abgearbeitet.

Syntax: (Anweisung)

$$
\begin{array}{lcl}
<\text{Anweisung}> & ::= & <\text{EinfacheAnw}> \mid \underline{\{}\ \{\ <\text{EinfacheAnw}>\ \}^{+}\ \underline{\}} \\
<\text{EinfacheAnw}> & ::= & <\text{VarDef}>\ \underline{;} \mid <\text{Zuweisung}>\ \underline{;} \mid \\
 & & <\text{Selektion}> \mid \ldots
\end{array}
$$

Beispiel

Die folgende Funktion berechnet `fib(4)`. `b` enthält die letzte und `a` die vorletzte Fibonaccizahl.

```
int f4()
{
  int a = 0;        // a = fib(0)
  int b = 1;        // b = fib(1)
  int t;

  t = a+b; a = b; b = t;     // b = fib(2)
  t = a+b; a = b; b = t;     // b = fib(3)
  t = a+b; a = b; b = t;     // b = fib(4)
  return b;
}
```

Bemerkung: Die Variable `t` wird benötigt, da die beiden Zuweisungen

$$\left\{ \begin{array}{lcl} \texttt{b} & \leftarrow & \texttt{a+b} \\ \texttt{a} & \leftarrow & \texttt{b} \end{array} \right\}$$

nicht gleichzeitig durchgeführt werden können.

Bemerkung: Man beachte, dass die Reihenfolge in

```
t = a+b;
a = b;
b = t;
```

nicht vertauscht werden darf. In der funktionalen Programmierung mussten wir hingegen weder auf die Reihenfolge achten noch irgendwelche „Hilfsvariablen“ einführen.

Bedingte Anweisung (Selektion)

Anstelle des `cond`-Operators wird in der imperativen Programmierung die bedingte Anweisung verwendet.

Syntax: (Bedingte Anweisung, Selektion)

$$\begin{aligned} <\text{Selektion}> \quad ::= \quad & \underline{\text{if}}\ (\ <\text{BoolAusdr}>\ )\ <\text{Anweisung}> \\ & [\ \underline{\text{else}}\ <\text{Anweisung}>\ ] \end{aligned}$$

Ist die Bedingung in runden Klammern wahr, so wird die erste Anweisung ausgeführt, ansonsten die zweite Anweisung nach dem `else` (falls vorhanden).

Genauer bezeichnet man die Variante **ohne** `else` als bedingte Anweisung, die Variante **mit** `else` als Selektion.

Beispiel: Die funktionale Form

```
int absolut( int x )
{
  return cond( x<=0, -x, x );
}
```

ist äquivalent zu

```
int absolut( int x )
{
  if ( x <= 0 )
    return -x;
  else
    return x;
}
```

While-Schleife

Iterative Prozesse sind so häufig, dass man hierfür eine Abstraktion schaffen will. In C++ gibt es dafür verschiedene imperative Konstrukte, die wir jetzt kennenlernen.

Programm: (Fakultät mit While-Schleife [fakwhile.cc])

```
int fak( int n )
{
  int ergebnis = 1;
  int zaehler = 2;

  while ( zaehler <= n )
  {
    ergebnis = zaehler*ergebnis;
    zaehler  = zaehler+1;
  }
  return ergebnis;
}
```

Syntax: **(While-Schleife)**

$$<\text{WhileSchleife}> \quad ::= \quad \underline{\text{while}}\ \underline{(}\ <\text{BoolAusdr}>\ \underline{)}\ <\text{Anweisung}>$$

Die Anweisung wird solange ausgeführt wie die Bedingung erfüllt ist.

Wir überlegen informell warum das Beispiel funktioniert.

- Ist $n = 0$ oder $n = 1$, also $n < 2$ (andere Zahlen sind nicht erlaubt), so ist die Schleifenbedingung nie erfüllt und das Ergebnis ist 1, was korrekt ist.

- Ist $n \geq 2$ so wird die Schleife mindestens einmal durchlaufen. In jedem Durchlauf wird der zaehler dranmultipliziert und dann erhöht. Es werden also sukzessive die Zahlen $2, 3, \ldots$ an den aktuellen Wert multipliziert. Irgendwann erreicht zaehler den Wert n und damit ergebnis den Wert $n!$. Da zaehler nun den Wert $n + 1$ hat, wird die Schleife verlassen.

Später werden wir eine formale Methode kennenlernen, mit der man beweisen kann, dass das Programm korrekt funktioniert.

For-Schleife

Die obige Anwendung der `while`-Schleife ist ein Spezialfall, der so häufig vorkommt, dass es dafür eine Abkürzung gibt:

Syntax: (For-Schleife)

$$
\begin{array}{lll}
\langle \text{ForSchleife} \rangle & ::= & \underline{\text{for}}\ \underline{(}\ \langle \text{Init} \rangle\ \underline{;}\ \langle \text{BoolAusdr} \rangle\ \underline{;}\ \langle \text{Increment} \rangle\ \underline{)} \\
 & & \quad \langle \text{Anweisung} \rangle \\
\langle \text{Init} \rangle & ::= & \langle \text{VarDef} \rangle \mid \langle \text{Zuweisung} \rangle \\
\langle \text{Increment} \rangle & ::= & \langle \text{Zuweisung} \rangle
\end{array}
$$

Init entspricht der Initialisierung des Zählers, BoolAusdr der Ausführungsbedingung und Increment der Inkrementierung des Zählers.

Programm: **(Fakultät mit For-Schleife [fakfor.cc])**

```
int fak( int n )
{
  int ergebnis = 1;

  for ( int zaehler=2; zaehler<=n; zaehler=zaehler+1 )
  {
    ergebnis = zaehler*ergebnis;
  }

  return ergebnis;
}
```

Bemerkung:

- Eine `For`-Schleife kann direkt in eine `While`-Schleife transformiert werden. Dabei wird die Laufvariable vor der Schleife definiert.

- Der Gültigkeitsbereich von `zaehler` erstreckt sich nur über die `for`-Schleife (ansonsten hätte man es wie `ergebnis` außerhalb der Schleife definieren müssen). Wir werden später sehen wie man den Gültigkeitsbereich gezielt mit neuen Umgebungen kontrollieren kann.

- Die Initialisierungsanweisung enthält Variablendefinition und Initialisierung.

- Wie beim Fakultätsprogramm mit `while` wird die Inkrementanweisung am Ende des Schleifendurchlaufes ausgeführt.

Wiederholung

Lokale Konstanten und Variablen:

```
int f( int x )
{
  const int a = 3; // Konstante, kann nicht geaendert werden

  int b = x;       // Variablendefinition mit Initialisierung

  int c;           // Variablendefinition ohne Initialisierung
}
```

Zuweisung:

```
c = 3 * a + f(b); // Variable wird Wert des Ausdrucks " zugewiesen "
```

Auswertung eines Ausdrucks erfolgt relativ zu einer Umgebung, die eine Bindungstabelle enthält.

Anweisungen und Anweisungsfolgen:

- Anweisungen werden immer mit einem Semikolon beendet.
- Anweisungsfolgen werden der Reihe nach bearbeitet und in { } gesetzt.
- Wo eine einzelne Anweisung stehen kann darf auch eine Folge von Anweisungen stehen.

`if`-Anweisung (Bedingte Anweisung, Selektion):

```
if ( a < 0 ) b = f(c); else b = f(3);
if ( a < 0 ) { b = f(c); c = 0; } else { b = f(3); c = 5; }
```

`while`-Schleife:

```
int i = 10; while ( i >= 0 ) i = i-1;
int i = 10; int b = 0; while ( i >= 0 ) { b = b+f(i); i = i-1; }
```

`for`-Schleife:

```
for ( int i=0; i>=0; i=i-1 ) ;
int b = 0; for ( int i=0; i>=0; i=i-1 ) b = b+f(i);
```

Beispiele

Wir benutzen nun die neuen Konstruktionselemente um die iterativen Prozesse zur Berechnung der Fibonaccizahlen und der Wurzelberechnung nochmal zu formulieren.

Programm: **(Fibonacci mit For-Schleife [fibfor .cc])**

```
int fib( int n )
{
  int a = 0;
  int b = 1;
  for ( int i=0; i<n; i=i+1 )
  {
    int t = a+b; a = b; b = t;
  }
  return a;
}
```

Programm: **(Newton mit While-Schleife [newtonwhile.cc])**

```
#include "fcpp.hh"

double wurzel( double a )
{
  double x = 1.0;

  while ( fabs(x*x - a) > 1e-12 )
    x = 0.5 * (x + a/x);
  return x;
}

int main()
{
  return print( wurzel(2.0) );
}
```

Goto

Neben den oben eingeführten Schleifen gibt es eine alternative Möglichkeit die Wiederholung zu formulieren. Wir betrachten nochmal die Berechnung der Fakultät mittels einer `while`-Schleife:

```
int t = 1;
int i = 2;

while ( i <= n )
{
   t = t*i;
   i = i+1;
}
```

Mit der goto-Anweisung kann man den Programmverlauf an einer anderen, vorher markierten Stelle fortsetzen:

```
int t = 1; int i = 2;
anfang: if ( i > n ) return t;
t = t*i;
i = i+1;
goto anfang;
```

- anfang nennt man eine Sprungmarke (engl.: label). Jede Anweisung kann mit einer Sprungmarke versehen werden.
- Der Sprung kann nur innerhalb einer Funktion erfolgen.
- `While`- und `For`-Schleife können mittels `goto` und Selektion realisiert werden.

In einem berühmten Letter to the Editor [*Go To Statement Considered Harmful, Communications of the ACM, Vol. II, Number 3, 1968*] hat Edsger W. Dikstra[13] dargelegt, dass `goto` zu sehr unübersichtlichen Programmen führt und nicht verwendet werden sollte.

Man kann zeigen, dass `goto` nicht notwendig ist und man mit den obigen Schleifenkonstrukten auskommen kann. Dies nennt man strukturierte Programmierung. Die Verwendung von `goto` in C/C++ Programmen gilt daher als verpönt und schlechter Programmierstil!

Eine abgemilderte Form des `goto` stellen die **break**- und **continue**-Anweisung dar. Diese erhöhen, mit Vorsicht eingesetzt, die Übersichtlichkeit von Programmen.

[13] Edsger Wybe Dijkstra, 1930–2002, niederländischer Informatiker.

Regeln guter Programmierung

1. Einrückung sollte verwendet werden um Schachtelung von Schleifen bzw. `if`-Anweisungen anzuzeigen:

```
if ( x >= 0 )
{
  if ( y <= x )
    b = x - y; // b ist groesser 0
  else
  {
    while ( y > x )
      y = y - 1;
    b = x + y;
  }
  i = f(b);
}
```

2. Verwende möglichst sprechende Variablennamen! Kurze Variablennamen wie i, j, k sollten nur innerhalb (kurzer) Schleifen verwendet werden (oder wenn sie der mathematischen Notation entsprechen).

3. Beschränke die Gültigkeit von Konstanten und Variablen auf den kleinstem möglichen Bereich.

4. Nicht mit Kommentaren sparen! Wichtige Anweisungen oder Programmzweige sollten dokumentiert werden. Beim „Programmieren im Großen“ ist die Programmdokumentation natürlich ein wesentlicher Bestandteil der Programmierung.

5. Verletzung dieser Regeln werden wir in den Übungen ab sofort mit Punktabzug belegen!

6. To be continued . . .

Formale Programmverifikation

Das Verständnis selbst einfacher imperativer Programme bereitet einige Mühe. Übung und Erfahrung helfen hier zwar, aber trotzdem bleibt der Wunsch formal beweisen zu können, dass ein Programm „funktioniert“. Dies gilt insbesondere für sicherheitsrelevante Programme.

Eine solche „formale Programmverifikation“ erfordert folgende Schritte:

1. Eine formale Beschreibung dessen was das Programm leisten soll. Dies bezeichnet man als Spezifikation.

2. Einen Beweis dass das Programm die Spezifikation erfüllt.

3. Dies erfordert eine formale Definition der Semantik der Programmiersprache.

Beginnen wir mit dem letzten Punkt. Hier haben sich unterschiedliche Vorgehensweisen herausgebildet, die wir kurz beschreiben wollen:

- Operationelle Semantik. Definiere eine Maschine, die direkt die Anweisungen der Programmiersprache verarbeitet.

- Denotationelle Semantik. Beschreibe Wirkung der Anweisungen der Programmiersprache als Zustandsänderung auf den Variablen:

  – $v_1, \ldots, v_m$ seien die Variablen im Programm. $W(v_i)$ der Wertebereich von v_i.
  – $Z = W(v_1) \times \ldots \times W(v_m)$ ist die Menge aller möglichen Zustände.
  – Sei a eine Anweisung, dann beschreibt $F_a : Z \to Z$ die Wirkung der Anweisung.

- Axiomatische Semantik. Beschreibe Wirkung der Anweisungen mittels prädikatenlogischer Formeln. Man schreibt

$$P \ \{a\} \ Q$$

wobei P, Q Abbildungen in die Menge $\{\text{wahr}, \text{falsch}\}$, sog. Prädikate, und a eine Anweisung ist.

$P \ \{a\} \ Q$ bedeutet dann:

  - Wenn P **vor** der Ausführung von a wahr ist, dann gilt Q **nach** der Ausführung von a (P impliziert Q).
  - P heißt auch **Vorbedingung** und Q **Nachbedingung**.

Beispiel:

$$-1000 < x \leq 0 \ \{x = x - 1\} \ -1000 \leq x < 0$$

Der oben beschriebene Formalismus der axiomatischen Semantik heisst auch Hoare[14]-Kalkül. Für die gängigen Konstrukte imperativer Programmiersprachen, wie Zuweisung, Sequenz und Selektion, lassen sich Zusammenhänge zwischen Vor- und Nachbedingung herleiten.

Sei nun S ein Programmfragment oder gar das ganze Programm. Dann lassen sich mit dem Hoare-Kalkül entsprechende P und Q finden so dass

$$P \; \{S\} \; Q.$$

Schließlich lässt sich die Spezifikation des Programms ebenfalls mittels prädikatenlogischer Formeln ausdrücken. P_{SPEC} bezeichnet entsprechend die Vorbedingung (Bedingung an die Eingabe) unter der das Ergebnis Q_{SPEC} berechnet wird.

[14] Sir Charles Anthony Richard Hoare, geb. 1934, brit. Informatiker.

Für ein gegebenes Programm S, welches die Spezifikation implementieren soll, sei nun $P_{\mathsf{PROG}}\ \{S\}\ Q_{\mathsf{PROG}}$ gezeigt. Der formale Prozess der Programmverifikation besteht dann im folgenden Nachweis:

$$(P_{\mathsf{SPEC}} \Rightarrow P_{\mathsf{PROG}}) \wedge (P_{\mathsf{PROG}}\ \{S\}\ Q_{\mathsf{PROG}}) \wedge (Q_{\mathsf{PROG}} \Rightarrow Q_{\mathsf{SPEC}}). \qquad (1)$$

Dabei kann z. B. P_{SPEC} eine stärkere Bedingung als P_{PROG} sein. Beispiel:

$$-1000 < x \leq 0 \ \Rightarrow\ x \leq 0.$$

Der Nachweis von (1) liefert erst die **partielle Korrektheit**. Getrennt davon ist zu zeigen, dass das Programm terminiert. Kann man dies nachweisen ist der Beweis der **totalen Korrektheit** erbracht.

Der Nachweis von $P_{\mathsf{PROG}}\ \{S\}\ Q_{\mathsf{PROG}}$ kann durch automatische Theorembeweiser unterstützt werden.

Korrektheit von Schleifen mittels Schleifeninvariante

Wir betrachten nun eine Variante des Hoare-Kalküls, mit der sich die Korrektheit von Schleifenkonstrukten nachweisen lässt. Dazu betrachten wir eine `while`-Schleife in der kanonischen Form

while (B(v)) { v = H(v); }

mit

- $v = (v_1, \ldots, v_m)$ dem Vektor von Variablen, die im Rumpf modifiziert werden,
- $B(v)$, der Schleifenbedingung und
- $H(v) = (H_1(v_1, \ldots, v_m), \ldots, H_m(v_1, \ldots, v_m))$ dem **Schleifentransformator**.

Als Beispiel dient die Berechnung der Fakultät. Dort lautet die Schleife:

```
while ( i <= n ) { t = t*i; i = i+1; }
```

Also

$$v = (t, i) \qquad B(v) \equiv i \leq n \qquad H(v) = (t * i, i + 1).$$

Zusätzlich definieren wir die Abkürzung

$$H^j(v) = \underbrace{H(H(\ldots H(}_{j \text{ mal}} v) \ldots)).$$

Definition: (Schleifeninvariante)
Sei $v^j = H^j(v^0)$, $j \in \mathbb{N}_0$, die Belegung der Variablen nach j-maligem Durchlaufen der Schleife. Eine Schleifeninvariante INV(v) erfüllt:

1. INV(v^0) ist wahr.

2. INV$(v^j) \wedge B(v^j) \Rightarrow$ INV(v^{j+1}).

Gilt die Invariante vor Ausführung der Schleife und ist die Schleifenbedingung erfüllt, dann gilt die Invariante nach Ausführung des Schleifenrumpfes.

Angenommen, die Schleife wird nach k-maligem Durchlaufen verlassen, d. h. es gilt $\neg B(v^k)$. Ziel ist es nun zu zeigen, dass

$$\text{INV}(v^k) \wedge \neg B(v^k) \Rightarrow Q(v^k)$$

wobei $Q(v^k)$ die geforderte Nachbedingung ist.

Beispiel: Betrachte das Programm zur Berechnung der Fakultät von n:

```
t = 1; i = 2;
while ( i <= n ) { t = t*i; i = i+1; }
```

Behauptung: Sei $n \geq 1$, dann lautet die Schleifeninvariante:

$$\text{INV}(t, i) \equiv t = (i-1)! \wedge i - 1 \leq n.$$

1. Für $v^0 = (t^0, i^0) = (1, 2)$ gilt $\text{INV}(1, 2) \equiv 1 = (2-1)! \wedge (2-1) \leq n$. Wegen $n \geq 1$ ist das immer wahr.

2. Es gelte nun $\text{INV}(v^j) \equiv t^j = (i^j - 1)! \wedge i^j - 1 \leq n$ sowie $B(v^j) = i^j \leq n$. (Vorsicht v^j bedeutet **nicht** v hoch j!). Dann gilt

   - $t^{j+1} = t^j \cdot i^j = (i^j - 1)! \cdot i^j = i^j!$
   - $i^{j+1} = i^j + 1$, somit gilt wegen $i^j = i^{j+1} - 1$ auch $t^{j+1} = (i^{j+1} - 1)!$.
   - Schließlich folgt aus $B(i^j, t^j) \equiv i^j = i^{j+1} - 1 \leq n$ dass $\text{INV}(i^{j+1}, t^{j+1})$ wahr.

3. Am Schleifenende gilt $\neg(i \leq n)$, also $i > n$, also $i = n + 1$ da i immer um 1 erhöht wird. Damit gilt dann also

$$\begin{aligned} & \mathsf{INV}(i,t) \wedge \neg B(i,t) \\ \Leftrightarrow \quad & t = (i-1)! \wedge i - 1 \leq n \wedge i = n + 1 \\ \Leftrightarrow \quad & t = (i-1)! \wedge i - 1 = n \\ \Rightarrow \quad & t = n! \equiv Q(n) \end{aligned}$$

Für den Fall $n \geq 0$ muss man den Fall $0! = 1$ als Sonderfall hinzunehmen. Das Programm ist auch für diesen Fall korrekt und die Schleifeninvariante lautet $\mathsf{INV}(i,t) \equiv (t = (i-1)! \wedge i - 1 \leq n) \vee (n = 0 \wedge t = 1 \wedge i = 2)$.

Prozeduren und Funktionen

In der Mathematik ist eine Funktion eine Abbildung $f : X \to Y$. C++ erlaubt entsprechend die Definition n-stelliger Funktionen

```
int f( int x1, int x2 ) { return x1*x1 + x2; }
...
int y = f( 2, 3 );
```

In der Funktionalen Programmierung ist das einzig interessante an einer Funktion ihr Rückgabewert. Seiteneffekte spielen keine Rolle. In der Praxis ist das jedoch anders. Betrachte

```
void drucke( int x )
{
  int i = print( x ); // print aus fcpp.hh
}
...
```

```
drucke( 3 );
...
```

Es macht durchaus Sinn eine Funktion definieren zu können, deren einziger Zweck das Ausdrucken des Arguments ist. So eine Funktion hat keinen sinnvollen Rückgabewert, ihr einziger Zweck ist der Seiteneffekt.

C++ erlaubt dafür den Rückgabetyp **void** (nichts). Der Funktionsrumpf darf dann keine **return**-Anweisung enthalten, welche einen Wert zurückgibt.

Der Funktionsaufruf ist eine gültige Anweisung, allerdings ist dann keine Zuweisung des Rückgabewerts erlaubt (die Funktion gibt keinen Wert zurück).

Funktionen, die keine Werte zurückliefern heißen Prozeduren. In C++ werden Prozeduren durch den Rückgabetyp **void** gekennzeichnet.

C++ erlaubt auch die Verwendung von Funktionen als Prozeduren, d. h. man verwendet einfach den Rückgabewert **nicht**.

Benutzerdefinierte Datentypen

Die bisherigen Programme haben nur mit Zahlen (unterschiedlichen Typs) gearbeitet. „Richtige“ Programme bearbeiten allgemeinere Daten, z. B.

- Zuteilung der Studenten auf Übungsgruppen,
- Flugreservierungssystem,
- Textverarbeitungsprogramm, Zeichenprogramm, . . .

Bemerkung: Im Sinne der Berechenbarkeit ist das keine Einschränkung, denn auf beliebig großen Bändern (Turing-Maschine), in beliebig tief verschachtelten Funktionen (Lambda-Kalkül) oder in beliebig großen Zahlen (FC++ mit langen Zahlen) lassen sich beliebige Daten kodieren.
Da dies aber sehr umständlich und ineffizient ist, erlauben praktisch alle Programmiersprachen dem Programmierer die Definition neuer Datentypen.

Aufzählungstyp

Dies ist ein Datentyp, der aus endlich vielen Werten besteht. Jedem Wert ist ein Name zugeordnet.

Beispiel:

```
enum color { white, black, red, green, blue, yellow };
...
color bgcolor = white;
color fgcolor = black;
```

Syntax: **(Aufzählungstyp)**

<Enum> ::= enum <Identifikator>
{ <Identifikator> [, <Identifikator>] } ;

Programm: **(Vollständiges Beispiel [enum.cc])**

```
#include "fcpp.hh"

enum Zustand { neu, gebraucht, alt, kaputt };

void druckeZustand( Zustand x ) {
  if ( x == neu )       print( "neu" );
  if ( x == gebraucht ) print( "gebraucht" );
  if ( x == alt )       print( "alt" );
  if ( x == kaputt )    print( "kaputt" );
}

int main() { druckeZustand( alt ); }
```

Felder

Wir lernen nun einen ersten Mechanismus kennen, um aus einem bestehenden Datentyp, wie `int` oder `float`, einen neuen Datentyp zu erschaffen: das Feld (engl.: Array).

Definition: Ein Feld besteht aus einer $festen\ Anzahl$ von Elementen eines Grundtyps. Die Elemente sind angeordnet, d. h. mit einer Numerierung (Index) versehen. Die Numerierung ist fortlaufend und beginnt bei 0.

Bemerkung: In der Mathematik entspricht dies dem (kartesischen) Produkt von Mengen.

Beispiel: Das mathematische Objekt eines Vektors $x = (x_0, x_1, x_2)^T \in \mathbb{R}^3$ wird in C++ durch

```
double x[3];
```

dargestellt. Auf die Komponenten greift man wie folgt zu:

```
x[0] = 1.0;   // Zugriff auf das erste Feldelement
x[1] = x[0]; // das zweite
x[2] = x[1]; // und das letzte
```

D. h. die Größen `x[k]` verhalten sich wie jede andere Variable vom Typ `double`.

Syntax: (Felddefinition)

$$<\text{FeldDef}> \quad ::= \quad <\text{Typ}> <\text{Name: }> \underline{[} <\text{Anzahl}> \underline{]}$$

Erzeugt ein Feld mit dem Namen $<$Name: $>$, das $<$Anzahl$>$ Elemente des Typs $<$Typ$>$ enthält.

Bemerkung: Eine Felddefinition darf wie eine Variablendefinition verwendet werden.

Achtung: Bei der hier beschriebenen Felddefinition muss die Größe des Feldes zur **Übersetzungszeit bekannt sein**! Folgendes geht also **nicht**:

```
void f( int n )
{
  char s[n];
  ...
}
```

aber immerhin

```
const int n = 8;
char s[ 3*(n+1) ];
```

Vorsicht: Der GNU-C-Compiler erlaubt Felder variabler Größe als Spracherweiterung! Sie müssen die Optionen `-ansi -pedantic` verwenden, um für obiges Programm eine Fehlermeldung zu erhalten.

Sieb des Eratosthenes

Als Anwendung des Feldes betrachten wir eine Methode zur Erzeugung einer Liste von Primzahlen, die Sieb des Eratosthenes[15] genannt wird.

Idee: Wir nehmen eine Liste der natürlichen Zahlen größer 1 und streichen alle Vielfachen von $2, 3, 4, \ldots$. Alle Zahlen, die durch diesen Prozess $nicht$ erreicht werden, sind die gesuchten Primzahlen.

Bemerkung:

- Es genügt, nur die Vielfachen der Primzahlen zu nehmen (Primfaktorzerlegung).
- Um nachzuweisen, dass $N \in \mathbb{N}$ prim ist, reicht es, $k \nmid N$ (k ist kein Teiler von N) für alle Zahlen $k \in \mathbb{N}$ mit $k \leq \sqrt{N}$ zu testen.

[15] Eratosthenes von Kyrene, ca. 276–194 v. Chr., außergewöhnlich vielseitiger griechischer Gelehrter. Methode war damals schon bekannt, Eratosthenes hat nur den Begriff „Sieb“ geprägt.

Programm: **(Sieb des Eratosthenes [eratosthenes.cc])**

```
#include "fcpp.hh"

int main()
{
  const int n = 500000;
  bool prim[n];

  // Initialisierung
  prim[0] = false;
  prim[1] = false;
  for ( int i=2; i<n; i=i+1 )
    prim[i] = true;

  // Sieb
  for ( int i=2; i<=sqrt((double) n); i=i+1 )
```

```
    if ( prim[i] )
      for ( int j=2*i; j<n; j=j+i )
        prim[j] = false;

  // Ausgabe
  int m = 0;
  for ( int i=0; i<n; i=i+1 )
    if ( prim[i] )
      m = m+1;
  print( "Anzahl Primzahlen:" );
  print( m );

  return 0;
}
```

Bemerkung: Der Aufwand des Algorithmus lässt sich wie folgt abschätzen:

1. Der Aufwand der Initialisierung ist $\Theta(n)$.

2. Unter der Annahme einer „konstanten Primzahldichte“ erhalten wir

$$\text{Aufwand}(n) \leq C \sum_{k=2}^{\sqrt{n}} \frac{n}{k} = Cn \sum_{k=2}^{\sqrt{n}} \frac{1}{k} \leq Cn \int_1^{\sqrt{n}} \frac{dx}{x} = Cn \log \sqrt{n} = \frac{C}{2} n \log n$$

3. $O(n \log n)$ ist bereits eine fast optimale Abschätzung, da der Aufwand ja auch $\Omega(n)$ ist. Man kann die Ordnungsabschätzung daher nicht wesentlich verbessern, selbst wenn man zahlentheoretisches Wissen über die Primzahldichte hinzuziehen würde.

Bemerkung: Die Beziehung zur funktionalen Programmierung ist etwa folgende:

- Mit großen Feldern operierende Algorithmen kann man nur schlecht rein funktional darstellen.

- Dies ist vor allem eine Effizienzfrage, weil oft kleine Veränderungen großer Felder verlangt werden. Bei funktionaler Programmierung müsste man ein neues Feld erzeugen und als Rückgabewert verwenden.

- Algorithmen wie das Sieb des Eratosthenes formuliert man daher funktional auf andere Weise (Datenströme, Streams), was interessant ist, allerdings manchmal auch recht komplex wird. (Allerdings ist dieser Programmierstil auf neuen Prozessoren wie Grafikkarten interessant).

Zeichen und Zeichenketten

Datentyp `char`

- Zur Verarbeitung von einzelnen Zeichen gibt es den Datentyp `char`, der genau ein Zeichen aufnehmen kann:

```
char c = '%';
```

- Die Initialisierung benutzt die einfachen Anführungsstriche.

- Der Datentyp `char` ist kompatibel mit `int` (Zeichen entsprechen Zahlen im Bereich $-128 \ldots 127$). Man kann daher sogar mit ihm rechnen:

```
char c1 = 'a';
char c2 = 'b';
char c3 = c1+c2;
```

Normalerweise sollte man diese Eigenschaft aber nicht brauchen!

ASCII

Die den Zahlen $0 \ldots 127$ zugeordneten Zeichen nennt man den American Standard Code for Information Interchange oder kurz ASCII. Den druckbaren Zeichen entsprechen die Werte $32 \ldots 127$.

Programm: **(ASCII.cc)**

```
#include "fcpp.hh"

int main()
{
  for ( int i=32; i<=127; i=i+1 )
    print( i, (char) i, 0 );
}
```

Bemerkung:

- Das dritte Argument von `print` ist der (ignorierte) Rückgabewert.
- Die Zeichen $0, \ldots, 31$ dienen Steuerzwecken wie Zeilenende, Papiervorschub, Piepston, etc.
- Für die negativen Werte $-128, \ldots, -1$ (entspricht $128, \ldots, 255$ bei vorzeichenlosen Zahlen) gibt es verschiedene Belegungstabellen (ISO 8859-n), mit denen man Zeichensätze und Sonderzeichen anderer Sprachen abdeckt.
- Noch komplizierter wird die Situation, wenn man Zeichensätze für Sprachen mit sehr vielen Zeichen (Chinesisch, Japanisch, etc) benötigt, oder wenn man mehrere Sprachen gleichzeitig behandeln will.
  Stichwort: Unicode.

Zeichenketten

Zeichenketten realisiert man in C am einfachsten mittels einem `char`-Feld. Konstante Zeichenketten kann man mit doppelten Anführungsstrichen auch direkt im Programm eingeben.

Beispiel: Initialisierung eines `char`-Felds:

```
char c[10] = "Hallo";
```

Bemerkung: Das Feld muss groß genug sein, um die Zeichenkette samt einem Endezeichen (in C das Zeichen mit ASCII-Code 0) aufnehmen zu können. Diese feste Größe ist oft sehr unhandlich, und viele Sicherheitsprobleme entstehen aus der Verwendung von zu kurzen `char`-Feldern von unachtsamen C-Programmierern!

Programm: **(Zeichenketten im C-Stil [Cstring.cc])**

```
#include "fcpp.hh"

int main()
{
  char name[32] = "Alan Turing";

  for ( int i=0; name[i]!=0; i=i+1 )
    print( name[i] );          // einzelne Zeichen
  print( name );               // ganze Zeichenkette
}
```

In C++ gibt es einen Datentyp `string`, der sich besser zur Verarbeitung von Zeichenketten eignet als bloße `char`-Felder:

Programm: (Zeichenketten im C++-Stil [CCstring.cc])

```
#include "fcpp.hh"
#include <string>

int main()
{
   std::string vorname = "Alan";
   std::string nachname = "Turing";
   std::string name = vorname + " " + nachname;
   print( name );
}
```

Dies erfordert das einbinden der Header-Datei `string` mit dem **#include** Befehl.

Typedef

Mittels der `typedef`-Anweisung kann man einem bestehenden Datentyp einen neuen Namen geben.

Beispiel:

```
typedef int MyInteger;
```

Damit hat der Datentyp `int` auch den Namen `MyInteger` erhalten.

Bemerkung: `MyInteger` ist kein neuer Datentyp. Er darf synonym mit `int` verwendet werden:

```
MyInteger x = 4;    // ein MyInteger
int y = 3;          // ein int

x = y;              // Zuweisung OK, Typen identisch
```

Anwendung: Verschiedene Computerarchitekturen (Rechner/Compiler) verwenden unterschiedliche Größen etwa von `int`-Zahlen. Soll nun ein Programm portabel auf verschiedenen Architekturen laufen, so kann man es an kritischen Stellen mit `MyInteger` schreiben. `MyInteger` kann dann an einer Stelle architekturabhängig definiert werden.

Beispiel: Auch Feldtypen kann man einen neuen Namen geben:

```
typedef double Punkt3d[3];
```

Dann kann man bequem schreiben:

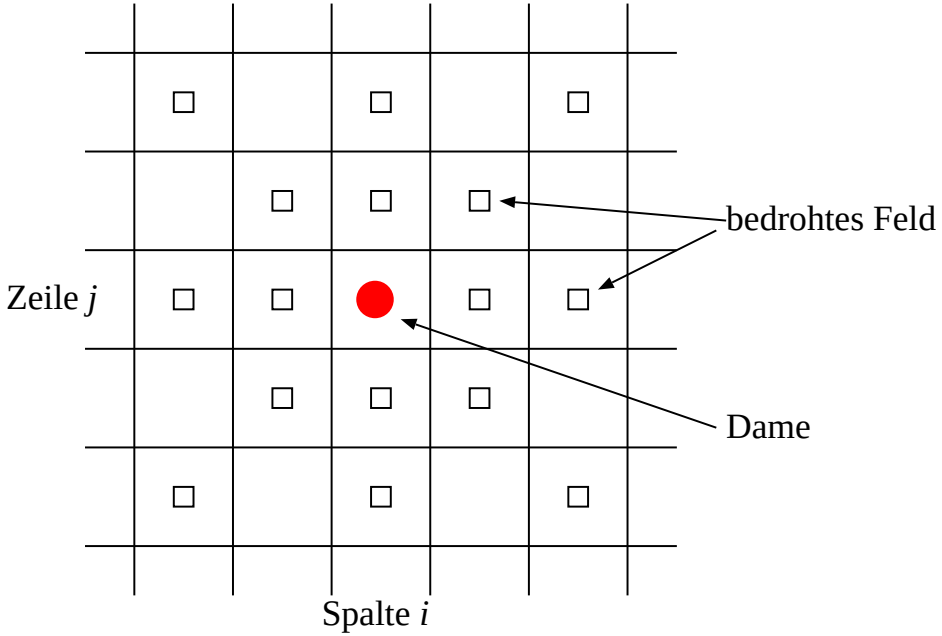
```
Punkt3d a, b;
a[0] = 0.0; a[1] = 1.0; a[2] = 2.0;
b[0] = 0.0; b[1] = 1.0; b[2] = 2.0;
```

Bemerkung: Ein Tipp zur Syntax: Man stelle sich eine Felddefinition vor und schreibt `typedef` davor.

Das Acht-Damen-Problem

Problem: Wie kann man acht Damen so auf einem Schachbrett positionieren, dass sie sich nicht gegenseitig schlagen können?

Zugmöglichkeiten der Dame: horizontal, vertikal, diagonal



Bemerkung:

- Ist daher die Dame an der Stelle (i, j), so bedroht sie alle (i', j') mit
  - $i = i'$ oder $j = j'$
  - $(i - i') = (j - j')$ oder $(i - i') = -(j - j')$

- Bei jeder Lösung steht in jeder Zeile/Spalte des Bretts genau eine Dame.

Idee:

- Man baut die Lösungen sukzessive auf, indem man erst in der ersten Zeile eine Dame platziert, dann in der zweiten, usw.

- Die Platzierung der ersten n Damen kann man durch ein `int`-Feld der Länge n beschreiben, wobei jede Komponente die Spaltenposition der Dame enthält.

Programm: (Acht-Damen-Problem [queens.cc])

```
#include "fcpp.hh"
#include <string>

const int board_size = 8;           // globale Konstante
typedef int columns[board_size];    // neuer Datentyp "columns"

bool good_position( int new_row, columns queen_cols, int new_col )
{
  for ( int row=0; row<new_row; row=row+1 )
    if ( ( queen_cols[row] == new_col ) ||
         ( new_row-row == abs( queen_cols[row]-new_col ) ) )
      return false;
  return true;
}

void display_board( columns queen_cols )
{
```

```
  for ( int r=0; r<board_size; r=r+1 )
  {
    std::string s( "" );
    for ( int c=0; c<board_size; c=c+1 )
      if ( c != queen_cols[r] )
        s = s + ".";
      else
        s = s + "D";
    print( s );
  }
  print( "␣" );
}

int queen_configs( int row, columns queen_cols )
{
  if ( row == board_size )
  {
    display_board( queen_cols );
    return 1;
```

```
  }
  else
  {
    int nr_configs = 0;
    for ( int col=0; col<board_size; col=col+1 )
      if ( good_position( row, queen_cols, col ) )
      {
        queen_cols[row] = col;
        nr_configs = nr_configs +
                     queen_configs( row+1, queen_cols );
      }
    return nr_configs;
  }
}

int main()
{
  columns queen_cols;
  print( "Anzahl Loesungen" );
```

```
    print( queen_configs( 0, queen_cols ) );
    return 0;
}
```

Bemerkung: Dieses Programm benutzt ein weiteres neues Element:

- Es wurde eine globale Konstante `board_size` verwendet. Diese kann innerhalb aller Funktionen benutzt werden.

- Auch die Typdefinition kann innerhalb aller Funktionen benutzt werden.

- Ein Feld wird als Argument an eine Funktion übergeben. Solche Argumente werden **nicht** kopiert (!). Dazu später mehr. Solange man Argumente nur liest oder nur neue Feldelemente beschreibt (wie hier) kann man sich auch vorstellen das Argument würde kopiert werden.

- Daher bis auf weiteres: Vorsicht bei der Benutzung von Feldern als Argumente von Funktionen!

Bemerkung:

- Dieses Verfahren des Ausprobierens verschiedener Möglichkeiten durch eine sogenannte Tiefensuche in einem Baum ist als Backtracking bekannt.

- Für $n = 8$ gibt es 92 Lösungen. Eine davon ist

```
. . . . . . . D
. . . D . . . .
D . . . . . . .
. . D . . . . .
. . . . . D . .
. D . . . . . .
. . . . . . D .
. . . . D . . .
```

Wiederholung

- Feld (array) als wesentliches Konstruktionselement für neue Datentypen
- Zeichenketten sind Felder über dem Typ **char**. C-Variante und C++-Variante.
- **typedef** zur Einführung neuer Namen für Datentypen (kein neuer Datentyp!)
- Acht-Damen-Problem als Beispiel für die Anwendung eines Felds:
  - Bisheriger Zustand der Berechnung wird in einem Feld zusammengefasst
  - Zustandsmenge wächst im Laufe der Rechnung

Zusammengesetzte Datentypen

Bei zusammengesetzten Datentypen kann man eine beliebige Anzahl möglicherweise verschiedener (sogar zusammengesetzter) Datentypen zu einem neuen Datentyp kombinieren. Diese Art Datentypen nennt man Strukturen.

Beispiel: Aus zwei `int`-Zahlen erhalten wir die Struktur `Rational`:

```
struct Rational {  // Schluesselwort struct
  int zaehler;     // eine Liste von
  int nenner;      // Variablendefinitionen
};                 // Semikolon nicht vergessen
```

Dieser Datentyp kann nun wie folgt verwendet werden

```
Rational p;          // Definition einer Variablen
p.zaehler = 3;       // Initialisierung der Komponenten
p.nenner = 4;
```

Syntax: **(Zusammengesetzter Datentyp)**

$$\begin{array}{lll} <\text{StructDef}> & ::= & \underline{\text{struct}} <\text{Name}> \underline{\{} \{ <\text{Komponente}> \underline{;} \}^{+} \underline{\}} \underline{;} \\ <\text{Komponente}> & ::= & <\text{VarDef}> \mid <\text{FeldDef}> \mid \ldots \end{array}$$

Eine Komponente ist entweder eine Variablendefinition ohne Initialisierung oder eine Felddefinition. Dabei kann der Typ der Komponente selbst zusammengesetzt sein.

Bemerkung: Strukturen sind ein Spezialfall von sehr viel mächtigeren Klassen, die später im OO-Teil behandelt werden.

Bemerkung: Im Gegensatz zu Feldern kann man mit Strukturen (zusammengesetzten Daten) gut funktional arbeiten, siehe etwa Abelson&Sussman: *Structure and Interpretation of Computer Programs*.

Anwendung: Rationale Zahlen

Programm: **(Rationale Zahlen, die erste [Rational2.cc])**

```
#include "fcpp.hh"

struct Rational {
  int zaehler;
  int nenner;
};

// Abstraktion: Konstruktor und Selektoren

Rational erzeuge_rat( int z, int n )
{
  Rational t;
  t.zaehler = z;
  t.nenner = n;
  return t;
```

```
}

int zaehler( Rational q )
{
  return q.zaehler;
}

int nenner( Rational q )
{
  return q.nenner;
}

// Arithmetische Operationen

Rational add_rat( Rational p, Rational q )
{
  return erzeuge_rat( zaehler(p)*nenner(q) +
                      zaehler(q)*nenner(p),
                      nenner(p)*nenner(q) );
```

```
}

Rational sub_rat( Rational p, Rational q )
{
  return erzeuge_rat( zaehler(p)*nenner(q) -
                      zaehler(q)*nenner(p),
                      nenner(p)*nenner(q) );
}

Rational mul_rat( Rational p, Rational q )
{
  return erzeuge_rat( zaehler(p)*zaehler(q),
                      nenner(p)*nenner(q) );
}

Rational div_rat( Rational p, Rational q )
{
  return erzeuge_rat( zaehler(p)*nenner(q),
                      nenner(p)*zaehler(q) );
```

```
}

void drucke_rat( Rational p )
{
  print( zaehler(p), "/", nenner(p), 0 );
}

int main()
{
  Rational p = erzeuge_rat( 3, 4 );
  Rational q = erzeuge_rat( 5, 3 );
  drucke_rat( p ); drucke_rat( q );

  // p*q + p - p*p
  Rational r = sub_rat( add_rat( mul_rat( p, q ), p ),
                        mul_rat( p, p ) );
  drucke_rat( r );
  return 0;
}
```

Bemerkung: Man beachte die Abstraktionsschicht, die wir durch den Konstruktor `erzeuge_rat` und die Selektoren `zaehler` und `nenner` eingeführt haben. Diese Schicht stellt die sogenannte Schnittstelle dar, über die unser Datentyp verwendet werden soll.

Problem: Noch ist keine Kürzung im Programm eingebaut. So liefert das obige Programm $1104/768$ anstatt $23/16$.

Abhilfe: Normalisierung im Konstruktor:

```
Rational erzeuge_rat( int z, int n )
{
  int g;
  Rational t;

  if ( n < 0 ) { n = -n; z = -z; }
  g = ggT( std::abs( z ), n );
  t.zaehler = z / g;
  t.nenner = n / g;
  return t;
}
```

Bemerkung: Ohne die Verwendung des Konstruktors hätten wir in allen arithmetischen Funktionen Änderungen durchführen müssen, um das Ergebnis in Normalform zu bringen.

Komplexe Zahlen

Analog lassen sich komplexe Zahlen einführen:

Programm: (Komplexe Zahlen, Version 1 [Complex2.cc])

```
#include "fcpp.hh"

struct Complex
{
  float real;
  float imag;
};

Complex erzeuge_complex( float re, float im )
{
  Complex t;
  t.real = re; t.imag = im;
  return t;
}
```

```
float real( Complex q ) { return q.real; }
float imag( Complex q ) { return q.imag; }

Complex add_complex( Complex p, Complex q )
{
  return erzeuge_complex( real(p) + real(q),
                          imag(p) + imag(q) );
}

// etc

void drucke_complex( Complex p )
{
  print( real(p), "+i*", imag(p), 0 );
}

int main()
{
```

```
    Complex p = erzeuge_complex( 3.0, 4.0 );
    Complex q = erzeuge_complex( 5.0, 3.0 );
    drucke_complex( p );
    drucke_complex( q );
    drucke_complex( add_complex( p, q ) );
}
```

Bemerkung: Hier wäre bei Verwendung der Funktionen `real` und `imag` zum Beispiel auch eine Änderung der internen Darstellung zu Betrag/Argument ohne Änderung der Schnittstelle möglich.

Gemischtzahlige Arithmetik

Problem: Was ist, wenn man mit komplexen und rationalen Zahlen gleichzeitig rechnen will?

Antwort: Eine Möglichkeit, die bereits die Sprache C bietet, ist die folgende:

1. Führe eine sogenannte variante Struktur (Schlüsselwort `union`) ein, die entweder eine rationale oder eine komplexe Zahl enthalten kann $\rightsquigarrow$ neuer Datentyp `Combination`

2. Füge auch eine Kennzeichnung hinzu, um was für eine Zahl (rational/komplex) es sich tatsächlich handelt $\rightsquigarrow$ neuer Datentyp `Mixed`.

3. Funktionen wie `add_mixed` prüfen die Kennzeichnung, konvertieren bei Bedarf und rufen dann `add_rat` bzw. `add_complex` auf.

```
struct Rational { int n; int d; };

struct Complex { float re; float im; };

union Combination   // entweder oder!
{
  Rational p;
  Complex c;
};

enum Kind { rational, complex };

struct Mixed          // gemischte Zahl
{
  Kind a;             // welche bist Du?
  Combination com;    // benutze je nach Art
};
```

Bemerkung: Diese Lösung hat etliche Probleme:

- Umständlich und unsicher (überschreiben)
- Das Hinzufügen weiterer Zahlentypen macht eine Änderung von bestehenden Funktionen nötig
- Typprüfungen zur Laufzeit $\rightarrow$ keine optimale Effizienz
- Speicherplatzbedarf der größten Komponente
- Hätten gerne: Infix-Notation mit unseren Zahlen

Bemerkung: Einige dieser Probleme werden wir mit den objektorientierten Erweiterungen von C++ vermeiden. Man sollte sich allerdings auch klar machen, dass das Problem von Arithmetik mit verschiedenen Zahltypen und eventuell auch verschiedenen Genauigkeiten tatsächlich extrem komplex ist. Eine vollkommene Lösung darf man daher nicht erwarten.

Globale Variablen und das Umgebungsmodell

Globale Variablen

Bisher: Funktionen haben kein Gedächtnis! Ruft man eine Funktion zweimal mit den selben Argumenten auf, so liefert sie auch dasselbe Ergebnis.

Grund:

- Berechnung einer Funktion hängt nur von ihren Parametern ab.
- Die lokale Umgebung bleibt zwischen Funktionsaufrufen nicht erhalten.

Das werden wir jetzt ändern!

Beispiel: Konto

Ein Konto kann man einrichten (mit einem Anfangskapital versehen), man kann abheben (mit negativem Betrag auch einzahlen), und man kann den Kontostand abfragen.

Programm: **(Konto [konto1.cc])**

```
#include "fcpp.hh"

int konto; // die GLOBALE Variable

void einrichten( int betrag )
{
  konto = betrag;
}

int kontostand()
```

```
{
  return konto;
}

int abheben( int betrag )
{
  konto = konto - betrag;
  return konto;
}

int main()
{
  einrichten( 100 );
  print( abheben( 25 ) );
  print( abheben( 25 ) );
  print( abheben( 25 ) );
}
```

Bemerkung:

- Die Variable `konto` ist außerhalb jeder Funktion definiert.
- Die Variable `konto` wird zu Beginn des Programmes erzeugt und nie mehr zerstört.
- Alle Funktionen können auf die Variable `konto` zugreifen. Man nennt sie daher eine globale Variable.
- Die Funktionen `einrichten`, `kontostand` und `abheben` stellen die **Schnittstelle** zur Bearbeitung des Kontos dar.

Frage: Oben haben wir eingeführt, dass Ausdrücke relativ zu einer Umgebung ausgeführt werden. In welcher Umgebung liegt `konto`?

Das Umgebungsmodell

Die Auswertung von Funktionen und Ausdrücken mit Hilfe von Umgebungen nennt man *Umgebungsmodell* (im Gegensatz zum Substitutionsmodell).

Definition: **(Umgebung)**

- Eine Umgebung enthält eine Bindungstabelle, d. h. eine Zuordnung von Namen zu Werten.

- Es kann beliebig viele Umgebungen geben. Umgebungen werden während des Programmlaufes implizit (automatisch) oder explizit (bewusst) erzeugt bzw. zerstört.

- Die Menge der Umgebungen bildet eine Baumstruktur. Die Wurzel dieses Baumes heißt globale Umgebung.

- Zu jedem Zeitpunkt des Programmablaufes gibt es eine aktuelle Umgebung. Die Auswertung von Ausdrücken erfolgt relativ zur aktuellen Umgebung.

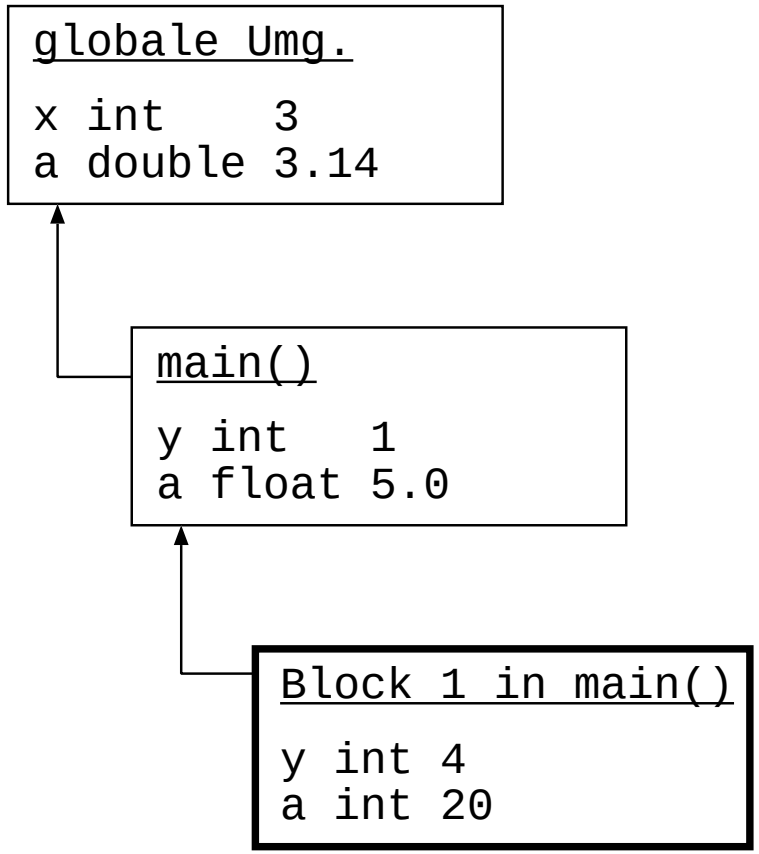
- Die Auswertung relativ zur aktuellen Umgebung versucht den Wert eines Namens in dieser Umgebung zu ermitteln, schlägt dies fehl, wird rekursiv in der nächst höheren („umschließenden“) Umgebung gesucht.

Eine Umgebung ist also relativ kompliziert. Das Umgebungsmodell beschreibt, wann Umgebungen erzeugt/zerstört werden und wann die Umgebung gewechselt wird.

Beispiel:

```
int x = 3;
double a = 4.3;          // 1
int main()
{
  int y = 1;
  float a = 5.0;         // 2
  {
    int y = 4;
    int a = 8;           // 3
    a = 5 * y;           // 4
    ::a = 3.14;          // 5
  }
}                        // 6
```

Nach Marke 5:



Eigenschaften:

- In einer Umgebung kann ein Name nur höchstens einmal vorkommen. In verschiedenen Umgebungen kann ein Name mehrmals vorkommen.

- Kommt auf dem Pfad von der aktuellen Umgebung zur Wurzel ein Name mehrmals vor, so verdeckt das erste Vorkommen die weiteren.

- Eine Zuweisung wirkt immer auf den sichtbaren Namen. Mit vorangestelltem `::` erreicht man die Namen der globalen Umgebung.

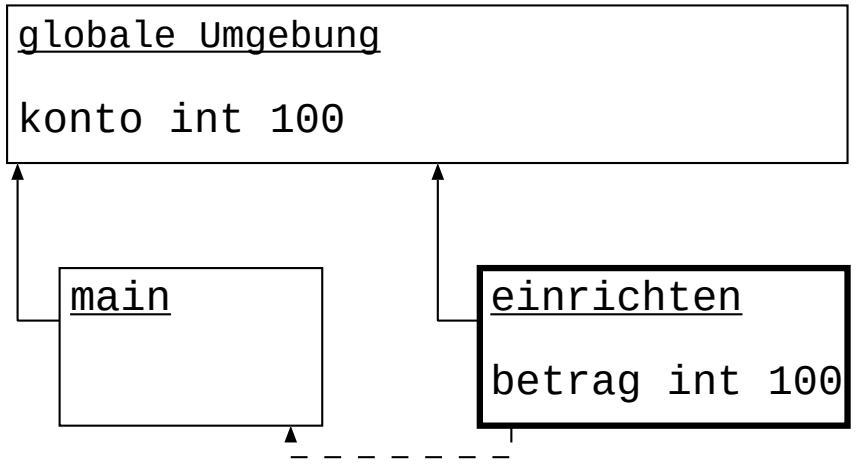
- Eine Anweisungsfolge in geschweiften Klammern bildet einen sogenannten Block, der eine eigene Umgebung besitzt.

- Eine Schleife `for ( int i=0; ...` wird in einer eigenen Umgebung ausgeführt. Diese Variable `i` gibt es im Rest der Funktion nicht.

Beispiel: (Funktionsaufruf)

```
int konto;

void einrichten( int betrag )
{
  konto = betrag;      // 2
}

int main()
{
  einrichten( 100 ); // 1
}
```

Nach Marke 2:



Bemerkung:

- Jeder Funktionsaufruf startet eine neue Umgebung unterhalb der globalen Umgebung. Dies ist dann die aktuelle Umgebung.

- Am Ende einer Funktion wird ihre Umgebung vernichtet und die aktuelle Umgebung wird die, in der der Aufruf stattfand (gestrichelte Linie).

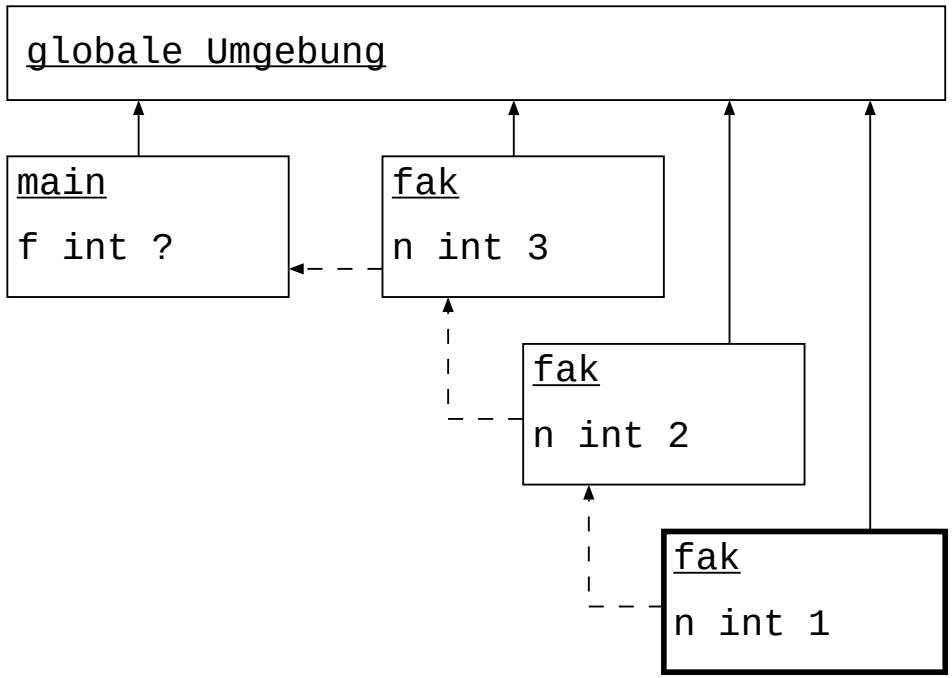
- Formale Parameter sind ganz normale Variablen, die mit dem Wert des aktuellen Parameters initialisiert sind.

- Sichtbarkeit von Namen ist in C++ am Programmtext abzulesen (statisch) und somit zur Übersetzungszeit bekannt. Sichtbar sind: Namen im aktuellen Block, nicht verdeckte Namen in umschließenden Blöcken und Namen in der globalen Umgebung.

Beispiel: (Rekursiver Aufruf)

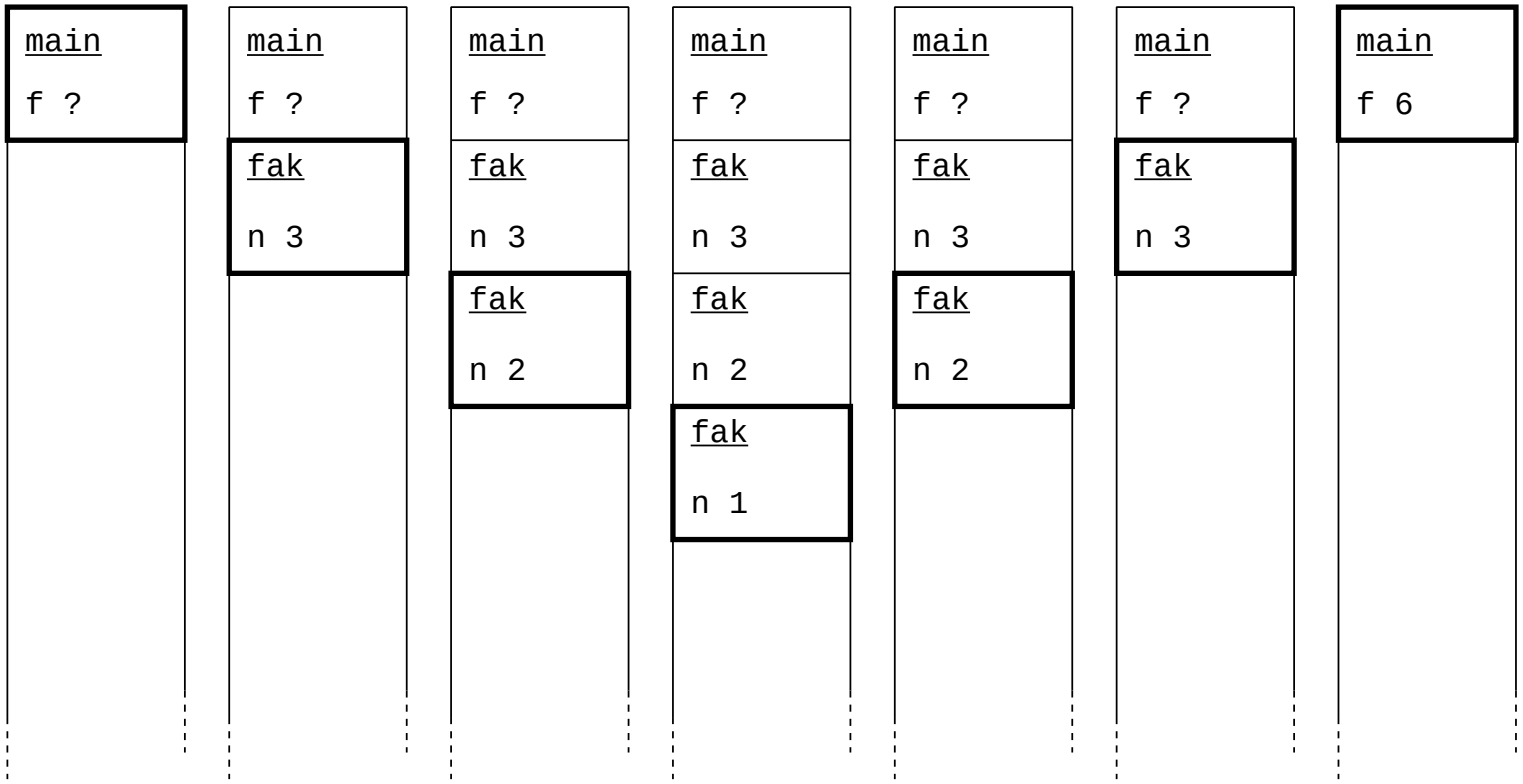
```
int fak( int n )
{
  if ( n == 1 )
    return 1;
  else
    return n * fak( n-1 );
}

int main()
{
  int f = fak( 3 ); // 1
}
```

Während Auswertung von `fak(3)`:



Bemerkung: Im obigen Beispiel gibt es zusätzlich noch eine „versteckte“ Variable für den Rückgabewert einer Funktion. Return kann als **Zuweisung über Umgebungsgrenzen** hinweg verstanden werden.

Beispiel: Die Berechnung von `fak(3)` führt zu:



Stapel

Bezeichnung: Die Datenstruktur, mit der das Umgebungsmodell normalerweise implementiert wird, nennt man Stapel, Stack oder LIFO (last in, first out). Im Deutschen ist, historisch bedingt, auch die Bezeichnung Keller gebräuchlich.

Definition: Ein Stapel ist eine Datenstruktur, welche folgende Operationen zur Verfügung stellt:

- Erzeugen eines leeren Stapels (*create*)
- Ablegen eines neuen Elements auf dem Stapel (*push*)
- Test, ob Stapel leer (*empty*)
- Holen des zuletzt abgelegten Elements vom Stapel (*pop*)

Programm: (Stapel [stack1.cc])

```
#include "fcpp.hh"

typedef int element_type;    // Integer-Stack

// START stack-library ...

const int stack_size = 1000;

element_type stack[stack_size];
int top = 0;  // Stapelzeiger (zeigt auf naechstes freies Element)

// Stack-Operationen

void push( element_type e )
{
  stack[top] = e;
  top = top + 1;
```

```
}

bool empty()
{
  return top == 0;
}

element_type pop()
{
  top = top - 1;
  return stack[top];
}

int main()
{
  push( 4 );
  push( 5 );
  while ( !empty() )
    print( pop() );
```

```
  return 0;
}
```

Bemerkung: Die Stapel-Struktur kann man verwenden, um rekursive in nicht rekursive Programme zu transformieren (wen es interessiert, findet unten eine nichtrekursive Variante für das Wechselgeld Beispiel). Dies ist aber normalerweise nicht von Vorteil, da der für Rekursion verwendete Stapel höchstwahrscheinlich effizienter verwaltet wird.

Programm: **(Wechselgeld nichtrekursiv [wg−stack.cc])**

```
#include "fcpp.hh"

int nennwert( int nr ) // Muenzart -> Muenzwert
{
  if ( nr == 1 ) return 1;  if ( nr == 2 ) return 2;
  if ( nr == 3 ) return 5;  if ( nr == 4 ) return 10;
  if ( nr == 5 ) return 50;
  return 0;
```

```
}

struct Arg              // Stapelelemente
{
  int betrag;           // das sind die Argumente der
  int muenzarten;       // rekursiven Variante
};

const int N = 1000;   // Stapelgroesse

int wechselgeld2( int betrag )
{
  Arg stapel[N];        // hier ist der Stapel
  int sp = 0;           // der "stack pointer"
  int anzahl = 0;       // das Ergebnis
  int b, m;             // Hilfsvariablen in Schleife

  stapel[sp].betrag = betrag;   // initialisiere St.
  stapel[sp].muenzarten = 5;    // Startwert
```

```
sp = sp + 1;                        // ein Element mehr

while ( sp > 0 )                    // Solange Stapel nicht leer
{
  sp = sp - 1;                      // lese oberstes Element
  b = stapel[sp].betrag;            // lese Argumente
  m = stapel[sp].muenzarten;

  if ( b == 0 )
    anzahl = anzahl + 1;            // Moeglichkeit gefunden
  else if ( b>0 && m>0 )
  {
    if ( sp >= N )
    {
      print( "Stapel zu klein" );
      return anzahl;
    }
    stapel[sp].betrag = b;              // Betrag b
    stapel[sp].muenzarten = m - 1;      // mit m - 1 Muenzarten
```

```
        sp = sp + 1;

        if ( sp >= N )
        {
          print( "Stapel zu klein" );
          return anzahl;
        }
        stapel[sp].betrag = b - nennwert(m);
        stapel[sp].muenzarten = m;      // mit m Muenzarten
        sp = sp + 1;
      }
    }

    return anzahl; // Stapel ist jetzt leer
}

int main()
{
    print( wechselgeld2( 300 ) );
```

```
}
```

Wiederholung

Zusammengesetzte Datentypen: **struct**, **union**

Erweiterung des Umgebungsmodels um globale Umgebung.

Globale Umgebung ist Wurzel einer Baumstruktur.

Funktionsaufruf startet neue Umgebung unterhalb der globalen Umgebung.

$\{...\}$ startet eine neue Umgebung unterhalb der aktuellen Umgebung.

Namen werden von der aktuellen Umgebung zur Wurzel hin gesucht. Globale Variable sind daher immer sichtbar, es sei denn, der Name ist verdeckt.

Umgebungen werden mittels eines Stapels gespeichert.

Umwandlung rekursiver in nichtrekursive Programme

Monte-Carlo Methode zur Bestimmung von π

Folgender Satz soll zur (näherungsweisen) Bestimmung von π herangezogen werden (randomisierter Algorithmus):

Satz: *Die Wahrscheinlichkeit q, dass zwei Zahlen $u, v \in \mathbb{N}$ keinen gemeinsamen Teiler haben, ist $\frac{6}{\pi^2}$. Zu dieser Aussage siehe [Knuth, Vol. 2, Theorem D].*

Um π zu approximieren, gehen wir daher wie folgt vor:

- Führe N „Experimente“ durch:
  - Ziehe „zufällig“ zwei Zahlen $1 \leq u_i, v_i \leq n$.
  - Berechne ggT(u_i, v_i).
  - Setze

$$e_i = \begin{cases} 1 & \text{falls ggT}(u_i, v_i) = 1 \\ 0 & \text{sonst} \end{cases}$$

- Berechne relative Häufigkeit $p(N) = \frac{\sum_{i=1}^{N} e_i}{N}$. Nach obigem Satz erwarten wir

$$\lim_{N \to \infty} p(N) = \frac{6}{\pi^2}.$$

- Also gilt $\pi \approx \sqrt{6/p(N)}$ für große N.

Pseudo-Zufallszahlen

Um Zufallszahlen zu erhalten, könnte man physikalische Phänomene heranziehen, von denen man überzeugt ist, dass sie „zufällig“ ablaufen (z. B. radioaktiver Zerfall). Solche Zufallszahl-Generatoren gibt es tatsächlich, sie sind allerdings recht teuer.

Daher begnügt man sich stattdessen oft mit Zahlenfolgen $x_k \in \mathbb{N}$, $0 \leq x_k < n$, welche deterministisch sind, aber zufällig „aussehen“. Für die „Zufälligkeit“ gibt es verschiedene Kriterien. Beispielsweise sollte jede Zahl gleich oft vorkommen, wenn man die Folge genügend lang macht:

$$\lim_{m\to\infty} \frac{|\{i|1 \leq i \leq m \wedge x_i = k\}|}{m} = \frac{1}{n}, \quad \forall k = 0, \ldots, n-1.$$

Einfachste Methode: **(Linear Congruential Method)** Ausgehend von einem x_0 verlangt man für $x_1, x_2, \ldots$ die Iterationsvorschrift

$$x_{n+1} = (ax_n + c) \bmod m.$$

Damit die Folge zufällig aussieht, müssen $a, c, m \in \mathbb{N}$ gewisse Bedingungen erfüllen, die man in [Knuth, Vol. 2, Kapitel 3] nachlesen kann.

Programm: (π mit Monte Carlo Methode [montecarlo1.cc])

```
#include "fcpp.hh"

int x = 93267;

unsigned int zufall()
{
  const int ia = 16807,  im = 2147483647;
  const int iq = 127773, ir = 2836;
  int k;

  k = x / iq;                          // LCG xneu = (a * xalt) mod m
  x = ia * ( x - k*iq ) - ir*k;        // a = 7^5, m = 2^31-1
  if ( x < 0 ) x = x + im;             // keine lange Arithmetik
  return x;                            // s. Numerical Recipes
}                                      // in C, Kap. 7.

unsigned int ggT( unsigned int a, unsigned int b )
```

```
{
  if ( b == 0 ) return a;
  else          return ggT( b, a % b );
}

int experiment()
{
  unsigned int x1, x2;

  x1 = zufall(); x2 = zufall();
  if ( ggT(x1, x2) == 1 )
    return 1;
  else
    return 0;
}

double montecarlo( int N )
{
  int erfolgreich = 0;
```

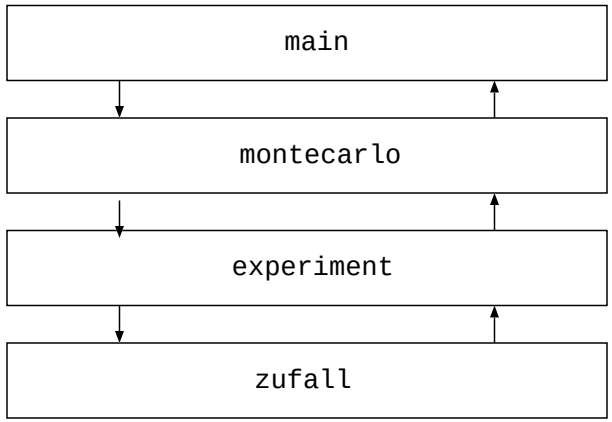
```
  for ( int i=0; i<N; i=i+1 )
    erfolgreich = erfolgreich + experiment();

  return ((double) erfolgreich) / ((double) N);
}

int main( int argc, char *argv[] )
{
  print( sqrt( 6.0/montecarlo( readarg_int( argc, argv, 1 ) ) ) );
}
```

Monte-Carlo funktional

Die Funktion `zufall` widerspricht offenbar dem funktionalen Paradigma (sonst müsste sie ja immer denselben Wert zurückliefern!). Stattdessen hat sie „Gedächtnis“ durch die globale Variable `x`.

Frage: Wie würde eine funktionale(re) Version des Programms ohne globale Variable aussehen?

Antwort: Eine Möglichkeit wäre es, `zufall` den Parameter x zu übergeben, woraus dann ein neuer Wert berechnet wird. Dieser Parameter müsste aber von `main` aus durch alle Funktionen hindurchgetunnelt werden:



Für `experiment`→`montecarlo` ist obendrein die Verwendung eines zusammengesetzten Datentyps als Rückgabewert nötig.

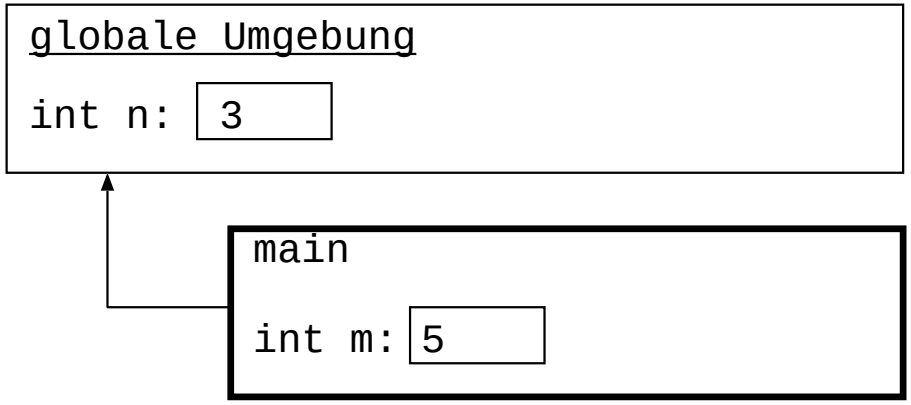
Beobachtung: In dieser Situation entstünde durch Beharren auf einem funktionalen Stil zwar kein Effizienzproblem, die Struktur des Programms würde aber deutlich komplizierter.

Zeiger und dynamische Datenstrukturen

Zeiger

Wir können uns eine Umgebung als Sammlung von Schubladen (Orten) vorstellen, die Werte aufnehmen können. Jede Schublade hat einen Namen, einen Typ und einen Wert:

```
int n = 3;

int main()
{
  int m = 5;
}
```



Idee: Es wäre nun praktisch, wenn man so etwas wie „Erhöhe den Wert in *dieser* Schublade (Variable) um eins“ ausdrücken könnte.

Anwendung: Im Konto-Beispiel möchten wir nicht nur ein Konto sondern viele Konten verwenden. Hierzu benötigt man einen Mechanismus, der einem auszu-

drücken erlaubt, $welches$ Konto verändert werden soll.

Idee: Man führt einen Datentyp Zeiger (Pointer) ein, der auf Variable (Schubladen) zeigen kann. Variablen, die Zeiger als Werte haben, heißen Zeigervariablen.

Bemerkung: Intern entspricht ein Zeiger der Adresse im physikalischen Speicher, an dem der Wert einer Variablen steht.

Notation: Die Definition „`int* x;`“ vereinbart, dass `x` auf Variablen (Schubladen) vom Typ `int` zeigen kann. Man sagt `x` habe den Typ „`int*`“.

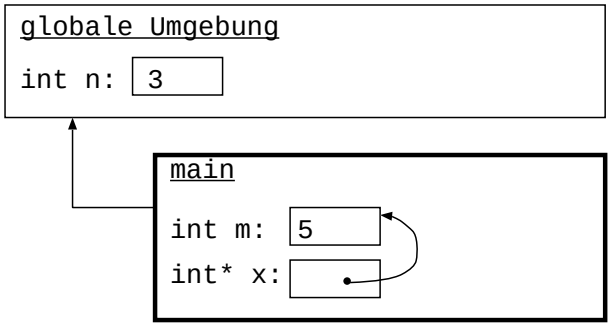
Die Zuweisung „`x = &n;`“ lässt `x` auf den Ort zeigen, an dem der Wert von `n` steht.

Die Zuweisung „`*x = 4;`“ verändert den Wert der Schublade, „auf die `x` zeigt“.

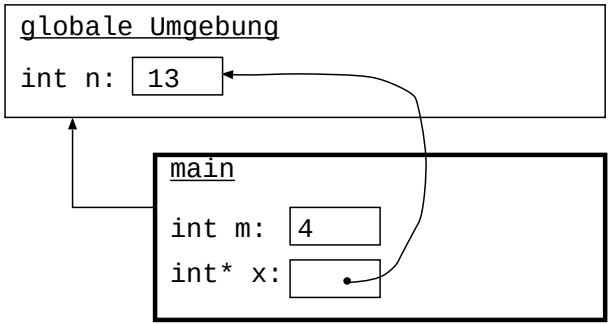
Beispiel:

```
int n = 3;

int main()
{
  int  m = 5;   // 1
  int* x = &m;  // 2
  *x = 4;       // 3
  x  = &n;      // 4
  *x = 13;      // 5
  return 0;
}
```

Nach (2)



Nach (5)



Zeiger im Umgebungsmodell

Im Umgebungsmodell gibt es eine Bindungstabelle, mittels derer jedem Namen ein Wert (und ein Typ) zugeordnet wird, etwa:

Name	Wert	Typ
`n`	3	`int`
`m`	5	`int`

Mathematisch entspricht das einer Abbildung w, die die *Symbole* $\mathtt{n}, \mathtt{m}$ auf die Wertemenge abbildet:

$$w : \{\mathtt{n}, \mathtt{m}\} \to \mathbb{Z}.$$

Die Zuweisung „`n = 3;`“ („=“ ist Zuweisung) manipuliert die Bindungstabelle so, dass nach der Zuweisung $w(\mathtt{n}) = 3$ („=“ ist Gleichheit) gilt.

Wenn auf der rechten Seite der Zuweisung auch ein Name steht, etwa „`n = m+1;`“, dann gilt nach der Zuweisung $w(\mathtt{n}) = w(\mathtt{m}) + 1$. Auf beide Namen wird also w angewandt.

Problem: Wir haben mehrere verschiedene Konten und möchten eine Funktion schreiben, die Beträge von Konten abhebt. In einer Variable soll dabei angegeben werden, von *welchem* Konto abgehoben wird.

Idee: Wir lassen Namen selbst wieder als Werte von (anderen) Namen zu, z. B.

Name	Wert	Typ
`n`	3	`int`
`m`	5	`int`
`x`	`n`	`int*`

Verwirklichung:

1. & ist der (einstellige) Adressoperator: „`x = &n`“ ändert die Bindungstabelle so, dass $w(\mathtt{x}) = \mathtt{n}$.

2. * ist der (einstellige) Dereferenzierungsoperator: Wenn `x = &n` gilt, so weist „`*x = 4;`“ dem Wert von n die Zahl 4 zu.

3. Den Typ eines Zeigers auf einen Datentyp `X` bezeichnet man mit „`X*`“.

Bemerkung:

- Auf der rechten Seite einer Zuweisung kann auf einen Namen der &-Operator genau einmal angewandt werden. Dieser $verhindert$ die Anwendung von w.

- Der $*$-Operator wendet die Abbildung w einmal auf das Argument rechts von ihm an. Der $*$-Operator kann mehrmals und sowohl auf der linken als auch auf der rechten Seite der Zuweisung angewandt werden.

Bemerkung: Auch eine Zeigervariable `x` kann wieder von einer anderen Zeigervariablen referenziert werden.

Diese hat dann den Typ „**int**∗∗“ oder „Zeiger auf eine **int**∗-Variable“.

Bemerkung: In C wird tendenziell die Notation „**int** ∗x;“ verwendet, wohingegen in C++ die Notation „**int**∗ x;“ empfohlen wird.

Allerdings impliziert die Notation „**int**∗ x, y;“ dass „x und y vom Typ **int**∗“ sind, was aber **nicht** zutrifft! (y ist vom Typ **int**)

Man liest „**int** ∗x“ als „x dereferenziert ist vom Typ **int**“ (in anderen Worten: „x zeigt auf **int**“).

Beispiel:

```
int    n = 3;
int    m = 5;
int*   x = &n;
int**  y = &x;
int*** z = &y;
```

Name	Wert	Typ
n	3	int
m	5	int
x	n	int*
y	x	int**
z	y	int***

Damit können wir schreiben

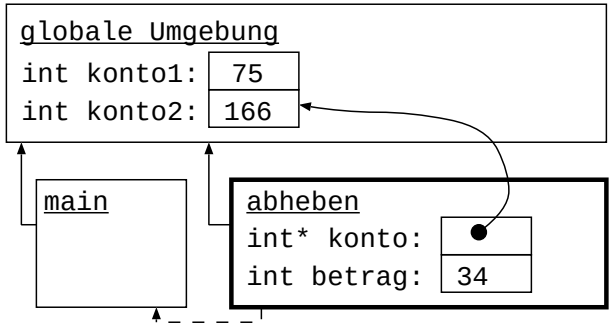
```
n = 4;     // das ist
*x = 4;    // alles
**y = 4;   // das
***z = 4; // gleiche !

x = &m;    // auch
*y = &m;   // das
**z = &m; // ist gleich !

y = &n;    // geht nicht, da n nicht vom Typ int*
y = &&n;   // geht auch nicht, da & nur
           // einmal angewandt werden kann
```

Call by reference

Programm: Die Realisation der Kontoverwaltung könnte wie folgt aussehen:

```
int konto1 = 100;
int konto2 = 200;

int abheben( int* konto, int betrag )
{
  *konto = *konto - betrag;      // 1
  return *konto;                 // 2
}

int main()
{
  abheben( &konto1, 25 );        // 3
  abheben( &konto2, 34 );        // 4
}
```

Nach Marke 1, im zweiten Aufruf von `abheben`:



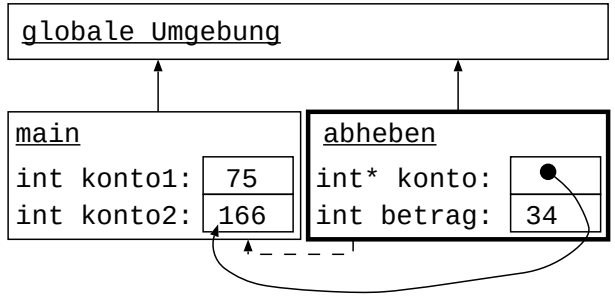
Definition: In der Funktion `abheben` nennt man `betrag` einen *call by value* Parameter und `konto` einen *call by reference* Parameter.

Bemerkung: Es gibt Computersprachen, die konsequent $call\ by\ value$ verwenden (z. B. Lisp/Scheme), und solche, die konsequent $call\ by\ reference$ verwenden (z. B. Fortran). Algol 60 war die erste Programmiersprache, die beides möglich machte.

Bemerkung: Die Variablen `konto1`, `konto2` im letzten Beispiel müssen nicht global sein! Folgendes ist auch möglich:

```
int abheben( int* konto, int betrag )
{
  *konto = *konto - betrag;   // 1
  return *konto;              // 2
}

int main()
{
  int konto1 = 100;
  int konto2 = 200;

  abheben( &konto1, 25 );     // 3
  abheben( &konto2, 34 );     // 4
}
```

Nach (1), zweiter Aufruf von abheben



Bemerkung: `abheben` darf `konto1` in `main` verändern, obwohl dieser Name dort

nicht sichtbar ist! Zeiger können also die Sichtbarkeitsregeln durchbrechen und — im Prinzip — kann somit auch jede lokale Variable von einer anderen Prozedur aus verändert werden.

Bemerkung: Es gibt im wesentlichen zwei Situationen in denen man Zeiger als Argumente von Funktionen einsetzt:

- Der Seiteneffekt ist explizit erwünscht wie in `abheben` ($\rightarrow$ Objektorientierung).
- Man möchte das Kopieren großer Objekte sparen ($\rightarrow$ `const` Zeiger).

Referenzen in C++

Beobachtung: Obige Verwendung von Zeigern als Prozedurparameter ist ziemlich umständlich: Im Funktionsaufruf müssen wir ein & vor das Argument setzen, innerhalb der Prozedur müssen wir den * benutzen.

Abhilfe: Wenn man in der Funktionsdefinition die Syntax **int**& x verwendet, so kann man beim Aufruf den Adressoperator & und bei der Verwendung innerhalb der Funktion den Dereferenzierungsoperator * weglassen. Dies ist wieder sogenannter „syntaktischer Zucker“.

Programm: **(Konto mit Referenzen)**

```
int abheben( int& konto, int betrag )
{
  konto = konto - betrag;   // 1
  return konto;             // 2
}
```

```
int main()
{
  int konto1 = 100;
  int konto2 = 200;

  abheben( konto1, 25 );   // 3
  abheben( konto2, 34 );   // 4
}
```

Bemerkung: Referenzen können nicht nur als Funktionsargumente benutzt werden:

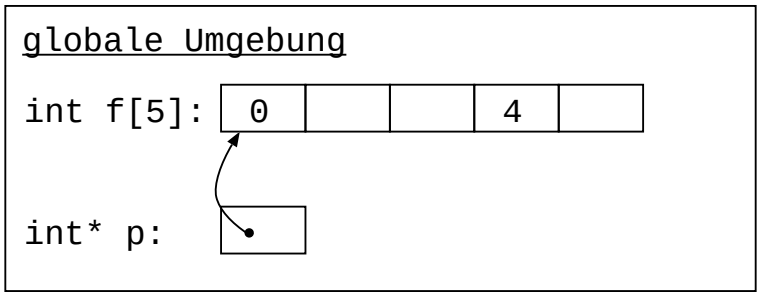
```
int  n = 3;
int& r = n; // independent reference
r = 5;      // selber Effekt n = 5;
```

Zeiger und Felder

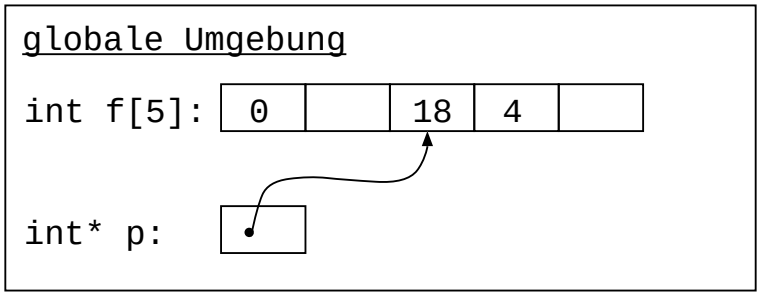
Beispiel: Zeiger und (eingebaute) Felder sind in C/C++ synonym:

```
int f[5];
int* p = f; // f hat Typ int*

...

p[0] = 0;
p[3] = 4;    // 1

p = &(f[2]);
*p = 18;     // p[0] = 18;
```

Nach Marke 1:



Am Ende



Bemerkung:

- Die Äquivalenz von eingebauten Feldern mit Zeigern ist eine höchst problematische Eigenschaft von C. Insbesondere führt sie dazu, dass man innerhalb einer mit einem Feld aufgerufenen Funktion die Länge dieses Feldes nicht direkt zur Verfügung hat. Diese muss anderweitig bekannt sein, oder es muss auf eine Bereichsüberprüfung verzichtet werden, was unter anderem ein Sicherheitsproblem darstellt (siehe auch die Diskussion bei `char`-Feldern).

- In C++ werden daher bessere Feldstrukturen (`vector`, `string`, `valarray`) in der Standard-Bibliothek STL (*Standard Template Library*) zur Verfügung gestellt.

Wiederholung

Zeigervariablen: Variablen, die auf Orte verweisen an denen der Wert einer Variablen gespeichert wird.

`int* p;` Zeiger auf eine `int` Variable.

`int n; p = &n; *p = 3;` Adresse-von- und Dereferenzierungsoperator.

`void f( int n, int* m );` Call-by-value und call-by-reference Parameter.

`void f( int n, int& m );` Referenzen in C++

`int a[10]; int* p = a; p[3] = 5;` (Eingebaute) Felder und Zeiger.

Zeiger erlauben die Änderungen beliebiger Werte von beliebigen Stellen innerhalb eines Programmes. Sie sollten mit großer Vorsicht eingesetzt werden.

Felder als Argumente von Funktionen

```
void f( int a[10] )
{
  a[3] = 17;
}
```

ist äquivalent zu

```
void f( int* a )
{
  a[3] = 17;
}
```

(Eingebaute) Felder werden also immer by reference übergeben!

Es findet in keinem Fall eine Bereichsprüfung statt und die Länge ist in f unbekannt.

Zeiger auf zusammengesetzte Datentypen

Beispiel: Es sind auch Zeiger auf Strukturen möglich. Ist `p` ein solcher Zeiger, so kann man mittels `p-><Komponente>` eine Komponente selektieren:

```
struct rational
{
  int n;
  int d;
};

int main()
{
  rational  q;
  rational* p = &q;
  (*p).n = 5;     // Zuweisung an Komponente n von q
  p->n = 5;       // eine Abkuerzung
}
```

Problematik von Zeigern

Beispiel: Betrachte folgendes Programm:

```
char* alphabet()
{
  char buffer[27];
  for ( int i=0; i<26; i++ ) buffer[i] = i + 65;
  buffer[26] = 0;
  return buffer;
}

int main()
{
  char* c = alphabet();
  print(c);
}
```

Beobachtung: Der Speicher für das lokale Feld ist schon freigegeben, aber den Zeiger darauf gibt es noch.

Bemerkung:

- Der gcc-Compiler warnt für das vorige Beispiel, dass ein Zeiger auf eine lokale Variable zurückgegeben wird. Er merkt allerdings schon nicht mehr, dass auch der Rückgabewert (Rückgabe-Adresse) `buffer+2` problematisch ist.

- Zeiger sind ein sehr *maschinennahes* Konzept (vgl. Neumann-Architektur). In vielen Programmiersprachen (z. B. Lisp, Java, etc.) sind sie daher für den Programmierer nicht sichtbar.

- Um die Verwendung von Zeigern sicher zu machen, muss man folgendes Prinzip beachten: Speicher darf nur dann freigegeben werden, wenn keine Referenzen darauf mehr existieren. Dies ist vor allem für die im nächsten Abschnitt diskutierte dynamische Speicherverwaltung wichtig.

Dynamische Speicherverwaltung

Bisher: Zwei Sorten von Variablen:

- Globale Variablen, die für die gesamte Laufzeit des Programmes existieren.
- Lokale Variablen, die nur für die Lebensdauer des Blockes/der Prozedur exisitieren.

Jetzt: Dynamische Variablen. Diese werden vom Programmierer explizit ausserhalb der globalen/aktuellen Umgebung erzeugt und vernichtet. Dazu dienen die Operatoren **new** und **delete**. Dynamische Variablen haben keinen Namen und können (in C/C++) nur indirekt über Zeiger bearbeitet werden.

Beispiel:

```
int m;
rational* p = new rational;
p->n = 4; p->d = 5;
m = p->n;
delete p;
```

Bemerkung:

- Die Anweisung `rational* p = new rational` erzeugt eine Variable vom Typ `rational` und weist deren Adresse dem Zeiger `p` zu. Man sagt auch, dass die Variable dynamisch allokiert wurde.

- Dynamische Variablen werden nicht auf dem Stack der globalen und lokalen Umgebungen gespeichert, sondern auf dem so genannten Heap. Dadurch ist es möglich, dass dynamisch allokierte Variablen in einer Funktion allokiert werden und die Funktion überdauern.

- Dynamische Variablen sind notwendig, um Strukturen im Rechner zu erzeugen, deren Größe sich während der Rechnung ergibt (und von der aufrufenden Funktion nicht gekannt wird).

- Die Größe der dynamisch allokierten Variablen ist nur durch den maximal verfügbaren Speicher begrenzt.

- Auch Felder können dynamisch erzeugt werden:

```
int   n = 18;
int* q = new int[n];   // Feld mit 18 int Eintraegen
q[5] = 3;
delete[] q;            // dynamisches Feld löschen
```

Probleme bei dynamischen Variablen

Beispiel: Wie schon im vorigen Abschnitt bemerkt, kann auf Zeiger zugegriffen werden, obwohl der Speicher schon freigegeben wurde:

```
int f()
{
   rational* p = new rational;
   p->n = 50;
   delete p;        // Vernichte Variable
   return p->n;     // Ooops, Zeiger gibt es immer noch
}
```

Beispiel: Wenn man alle Zeiger auf dynamisch allokierten Speicher löscht, kann dieser nicht mehr freigegeben werden ($\rightsquigarrow$ u.U. Speicherüberlauf):

```
int f()
{
   rational* p = new rational;
   p->n = 50;
   return p->n;   // Ooops, einziger Zeiger verloren
}
```

Problem: Es gibt zwei voneinander unabhängige Dinge, den Zeiger und die dynamische Variable. Beide müssen jedoch in konsistenter Weise verwendet werden. C++ stellt das nicht automatisch sicher!

Abhilfe:

- Manipulation der Variablen und Zeiger in Funktionen (später: Klassen) verpacken, die eine konsistente Behandlung sicherstellen.

- Benutzung spezieller Zeigerklassen (*smart pointers*).

- Die für den Programmierer angenehmste Möglichkeit ist die Verwendung von Garbage collection (= Sammeln von nicht mehr referenziertem Speicher).

Die einfach verkettete Liste

Zeiger und dynamische Speicherverwaltung benötigt man zur Erzeugung *dynamischer Datenstrukturen*.

Dies illustrieren wir am Beispiel der einfach verketteten Liste. Das komplette Programm befindet sich in der Datei `intlist.cc`.

Eine Liste natürlicher Zahlen

$$(12\ 43\ 456\ 7892\ 1\ 43\ 43\ 746)$$

zeichnet sich dadurch aus, dass

- die Reihenfolge der Elemente wesentlich ist, und
- Zahlen mehrfach vorkommen können.

Zur Verwaltung von Listen wollen wir folgende Operationen vorsehen

- Erzeugen einer leeren Liste.
- Einfügen von Elementen an beliebiger Stelle.
- Entfernen von Elementen.
- Durchsuchen der Liste.

Bemerkung: Der Hauptvorteil gegenüber dem Feld ist, dass das Einfügen und Löschen von Elementen schneller geschehen kann (es ist eine O(1)-Operation, wenn die Stelle nicht gesucht werden muss).

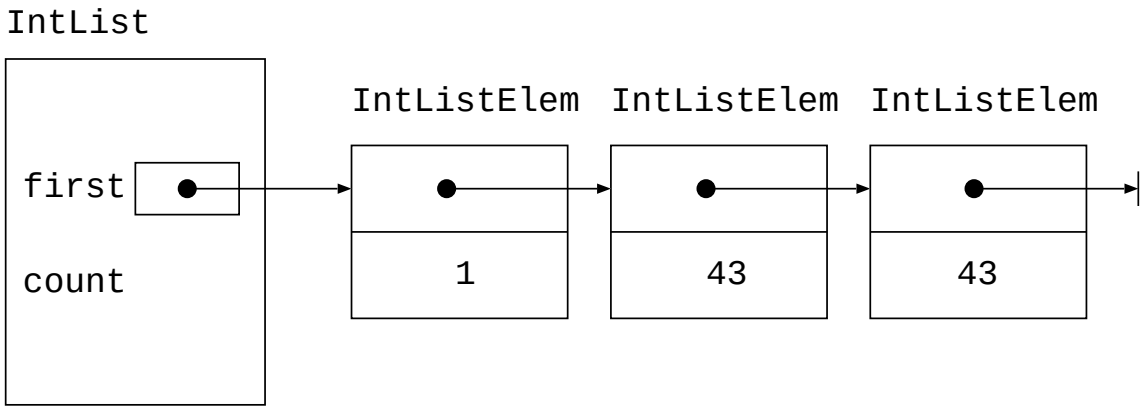
Eine übliche Methode zur Speicherung von Listen (natürlicher Zahlen) besteht darin ein $Listenelement$ zu definieren, das ein Element der Liste sowie einen Zeiger auf das nächste Listenelement enthält:

```
struct IntListElem
{
  IntListElem* next;   // Zeiger auf nächstes Element
  int value;           // Daten zu diesem Element
};
```

Um die Liste als Ganzes ansprechen zu können, definieren wir den folgenden zusammengesetzten Datentyp, der einen Zeiger auf das erste Element sowie die Anzahl der Elemente enthält:

```
struct IntList
{
  int count;           // Anzahl Elemente in der Liste
  IntListElem* first;  // Zeiger auf 1. Element der Liste
};
```

Das sieht also so aus:



Das Ende der Liste wird durch einen Zeiger mit dem Wert 0 gekennzeichnet.

Das klappt deswegen, weil 0 kein erlaubter Ort eines Listenelementes (irgendeiner Variable) ist.

Bemerkung: Die Bedeutung von 0 ist in C/C++ mehrfach überladen. In manchen Zusammenhängen bezeichnet es die Zahl 0, an anderen Stellen einen speziellen Zeiger. Auch der bool-Wert `false` ist synonym zu 0. In C++11 gibt es nun das Schlüsselwort nullptr .

Initialisierung

Folgende Funktion initialisiert eine `IntList`-Struktur mit einer leeren Liste:

```
void empty_list( IntList* l )
{
  l->first = 0;    // Liste ist leer
  l->count = 0;
}
```

Bemerkung: Die Liste wird *call-by-reference* übergeben, um die Komponenten ändern zu können.

Durchsuchen

Hat man eine solche Listenstruktur, so gelingt das Durchsuchen der Liste mittels

```
IntListElem* find_first_x( IntList l, int x )
{
  for ( IntListElem* p=l.first; p!=0; p=p->next )
    if ( p->value == x ) return p;
  return 0;
}
```

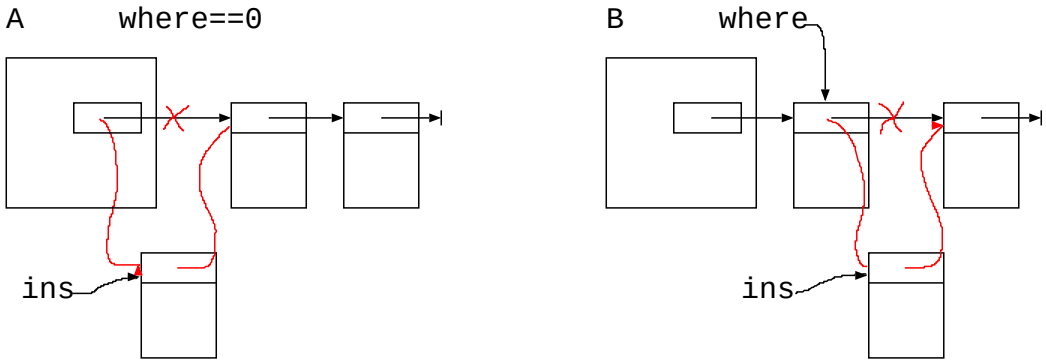
Einfügen

Beim Einfügen von Elementen unterscheiden wir zwei Fälle:

A Am Anfang der Liste einfügen.

B $Nach$ einem gegebenem Element einfügen.

Nur für diese beiden Fälle ist eine effiziente Realisierung der Einfügeoperation möglich. Es sind folgende Manipulationen der Zeiger erforderlich:



Programm: Folgende Funktion behandelt beide Fälle:

```
void insert_in_list( IntList* list, IntListElem* where,
                     IntListElem* ins )
{
  if ( where == 0 )
  { // fuege am Anfang ein
    ins->next = list->first;
    list->first = ins;
    list->count = list->count + 1;
  }
  else
  {
    // fuege nach where ein
    ins->next = where->next;
    where->next = ins;
    list->count = list->count + 1;
  }
```

```
}
```

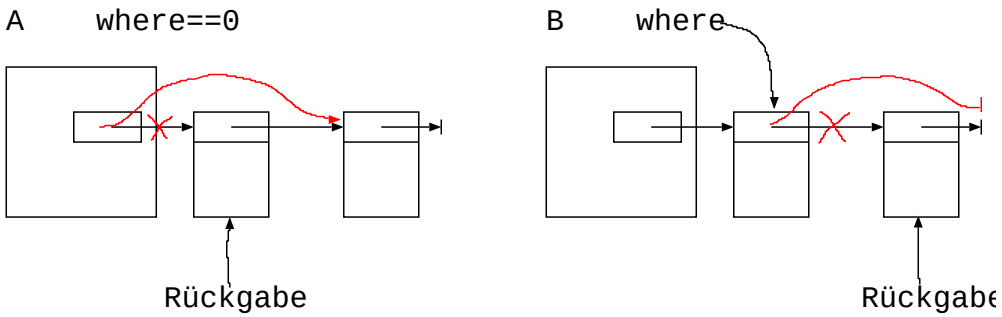
Entfernen

Auch beim Entfernen von Elementen unterscheiden wir wieder ob

1. das erste Element gelöscht werden soll, oder

2. das Element $nach$ einem gegebenem.

entsprechend graphisch:



Programm: Beide Fälle behandelt folgende Funktion:

```
IntListElem* remove_from_list( IntList* list ,
                               IntListElem* where )
{
  IntListElem* p;   // das entfernte Element

  // where == 0 dann entferne erstes Element
  if ( where == 0 ) {
    p = list->first;
    if ( p != 0 )
    {
      list->first = p->next;
      list->count = list->count - 1;
    }
    return p;
  }
```

```
    // entferne Element nach where
    p = where->next;
    if ( p != 0 ) {
      where->next = p->next;
      list->count = list->count - 1;
    }
    return p;
}
```

Bemerkung:

- Es wird angenommen, dass `where` ein Element der Liste ist. Wenn dies nicht erfüllt sein sollte, so ist Ärger garantiert!

- Alle Funktionen auf der Liste beinhalten *nicht* die Speicherverwaltung für die Objekte. Dies ist Aufgabe des benutzenden Programmteiles.

Kritik am Programmdesign

- Ob die Anzahl der Elemente überhaupt benötigt wird, hängt von der konkreten Anwendung ab. Man hätte die Liste auch einfach als Zeiger auf Listenelemente definieren können:

```
struct list_element
{
  list_element* next;        // Zeiger auf nächstes
  list_element_type value; // Datum dieses Elements
};
typedef list_element* list;
```

- Man wird auch Listen anderer Typen brauchen. Dies ist erst mit den später behandelten Werkzeugen wirklich befriedigend zu erreichen (Templates). Etwas mehr Flexibilität erhielte man aber über:

  ```
  typedef int list_element_type;
  ```

  und Verwendung dieses Datentyps später.

- Die Liste ist ein Spezialfall eines **Containers**. Mit der *Standard Template Library (STL)* bietet C++ eine leistungsfähige Implementierung von Containern unterschiedlicher Funktionalität.

Bemerkung: Lässt man nur das Einfügen und Löschen am **Listenanfang** zu, so implementiert die Liste das Verhalten eines Stacks mit dem Vorteil, dass man die maximale Zahl von Elementen nicht im Voraus kennen muss.

Listenvarianten

- Bei der doppelt verketteten Liste ist auch der Vorgänger erreichbar und die Liste kann auch in umgekehrter Richtung durchlaufen werden.

- Listen, die auch ein (schnelles) Einfügen am Ende erlauben, sind zur Implementation von Warteschlangen (Queues) nützlich.

- Manchmal kann man zirkuläre Listen gebrauchen. Diese sind für simple Speicherverwaltungsmechanismen (reference counting) problematisch.

- In dynamisch typisierten Sprachen können Elemente beliebigen Typ haben (zum Beispiel wieder Listen), und die Liste wird zum Spezialfall einer Baumstruktur. Am elegantesten ist dieses Konzept wohl in der Sprache Lisp ($= List\ Processing$) verwirklicht.

Wiederholung Zeiger

Verwendungsmuster für Zeiger:

- Call by reference
- Dynamisch allokierte Variablen
- Dynamische Datenstrukturen
  - Felder variabler Größe
  - Listen, Bäume, etc.

Problematik des Speicherüberlaufs $\rightarrow$ interagiert mit dem Betriebssystem:

- Exceptions
- Test auf nullptr

Endliche Menge

Im Gegensatz zu einer Liste kommt es bei einer endlichen Menge

$$\{34\ 567\ 43\ 1\}$$

1. nicht auf die Reihenfolge der Mitglieder an und

2. können Elemente auch nicht doppelt vorkommen!

Schnittstelle

Als Operationen auf einer Menge benötigen wir

- Erzeugen einer leeren Menge.
- Einfügen eines Elementes.
- Entfernen eines Elementes.
- Mitgliedschaft in der Menge testen.

In der nachfolgenden Implementierung einer Menge von **int**-Zahlen enthalten die genannten Funktionen auch die Speicherverwaltung!

Wir wollen zur Realisierung der Menge die eben vorgestellte einfach verkettete Liste verwenden.

Datentyp und Initialisierung

Datentyp: (Menge von Integer-Zahlen)

```
struct IntSet
{
  IntList list;
};
```

Man versteckt damit auch, dass `IntSet` mittels `IntList` realisiert ist.

Programm: (Leere Menge)

```
IntSet* empty_set()
{
  IntSet* s = new IntSet;
  empty_list( &s->list );
  return s;
```

```
}
```

Test auf Mitgliedschaft

Programm:

```
bool is_in_set( IntSet* s, int x )
{
  for ( IntListElem* p=s->list.first; p!=0; p=p->next )
    if ( p->value == x ) return true;
  return false;
}
```

Bemerkung:

- Dies nennt man sequentielle Suche. Der Aufwand ist $O(n)$ wenn die Liste n Elemente hat.
- Später werden wir bessere Datenstrukturen kennenlernen, mit denen man das in $O(\log n)$ Aufwand schafft (Suchbaum).

Einfügen in eine Menge

Idee: Man testet, ob das Element bereits in der Menge ist, ansonsten wird es am Anfang der Liste eingefügt.

```
void insert_in_set( IntSet* s, int x )
{
  if ( !is_in_set( s, x ) )
  {
    IntListElem* p = new IntListElem;
    p->value = x;
    insert_in_list( &s->list, 0, p );
  }
}
```

Bemerkung: Man beachte, dass diese Funktion auch die `IntListElem`-Objekte dynamisch erzeugt.

Ausgabe

Programm: (Ausgabe der Menge)

```
void print_set( IntSet* s )
{
  print( "{" );
  for ( IntListElem* p=s->list.first; p!=0; p=p->next )
    print( "␣␣", p->value, 0 );
  print( "}" );
}
```

Entfernen

Idee: Man sucht zuerst den Vorgänger des zu löschenden Elementes in der Liste und wendet dann die entsprechende Funktion für Listen an.

```
void remove_from_set( IntSet* s, int x )
{
  // Hat es ueberhaupt Elemente?
  if ( s->list.first == 0 ) return;

  // Teste erstes Element
  if ( s->list.first->value == x )
  {
    IntListElem* p = remove_from_list( &s->list, 0 );
    delete p;
    return;
  }

  // Suche in Liste, teste immer Nachfolger
```

```
  // des aktuellen Elementes
  for ( IntListElem* p=s->list.first; p->next!=0; p=p->next )
    if ( p->next->value == x )
    {
      IntListElem* q = remove_from_list( &s->list, p );
      delete q;
      return;
    }
}
```

Vollständiges Programm

Programm: (useintset.cc)

```
#include "fcpp.hh"
#include "intlist.cc"
#include "intset.cc"

int main()
{
  IntSet* s = empty_set();
  print_set( s );
  for ( int i=1; i<12; i=i+1 ) insert_in_set( s, i );
  print_set( s );
  for ( int i=2; i<30; i=i+2 ) remove_from_set( s, i );
  print_set( s );
}
```

Klassen

Motivation

Bisher:

- Funktionen bzw. Prozeduren (Funktion, bei welcher der Seiteneffekt wesentlich ist) als *aktive* Entitäten
- Daten als *passive* Entitäten.

Beispiel:

```
int konto1 = 100;
int konto2 = 200;
int abheben( int& konto, int betrag )
{
  konto = konto - betrag;
  return konto;
}
```

Kritik:

- Auf welchen Daten operiert `abheben`? Es könnte mit jeder **int**-Variablen arbeiten.
- Wir könnten `konto1` auch ohne die Funktion `abheben` manipulieren.
- Nirgends ist der Zusammenhang zwischen den globalen Variablen `konto1`, `konto2` und der Funktion `abheben` erkennbar.

Idee: Verbinde Daten und Funktionen zu einer Einheit!

Klassendefinition

Diese Verbindung von Daten und Funktionen wird durch Klassen (*classes*) realisiert:

Beispiel: Klasse für das Konto:

```
class Konto
{
public:
  int kontostand();
  int abheben( int betrag );
private:
  int k;
};
```

Sieht einer Definition eines zusammengesetzten Datentyps sehr ähnlich.

Syntax: **(Klassendefinition)** Die allgemeine Syntax der Klassendefinition lautet

$$<\text{Klasse}> \quad ::= \quad \underline{\text{class}} <\text{Name}> \underline{\{} <\text{Rumpf}> \underline{\}} \underline{;}$$

Im Rumpf werden sowohl Variablen als auch Funktionen aufgeführt. Bei den Funktionen genügt der Kopf. Die Funktionen einer Klasse heißen Methoden (*methods*). Alle Komponenten (Daten und Methoden) heißen Mitglieder. Die Daten heißen oft Datenmitglieder.

Bemerkung:

- Die Klassendefinition
  - beschreibt, aus welchen Daten eine Klasse besteht,
  - und welche Operationen auf diesen Daten ausgeführt werden können.
- Klassen sind (in C++) keine normalen Datenobjekte. Sie sind nur zur Kompilierungszeit bekannt und belegen daher keinen Speicherplatz.

Objektdefinition

Die Klasse kann man sich als Bauplan vorstellen. Nach diesem Bauplan werden Objekte (*objects*) erstellt, die dann im Rechner existieren. Objekte heißen auch Instanzen (*instances*) einer Klasse.

Objektdefinitionen sehen aus wie Variablendefinitionen, wobei die Klasse wie ein neuer Datentyp erscheint. Methoden werden wie Komponenten eines zusammengesetzten Datentyps selektiert und mit Argumenten wie eine Funktion versehen.

Beispiel:

```
Konto k1;
k1.abheben( 25 );

Konto* pk = &k1;
print( pk->kontostand() );
```

Bemerkung: Objekte haben einen internen Zustand, der durch die Datenmitglieder repräsentiert wird. Objekte haben ein Gedächtnis!

Kapselung

Der Rumpf einer Klassendefinition zerfällt in zwei Teile:

1. einen öffentlichen Teil, und

2. einen privaten Teil.

Der öffentliche Teil einer Klasse ist die Schnittstelle (*interface*) der Klasse zum restlichen Programm. Diese sollte für den Benutzer der Klasse ausreichende Funktionalität bereitstellen. Der private Teil der Klasse enthält Mitglieder, die zur Implementierung der Schnittstelle benutzt werden.

Bezeichnung: Diese Trennung nennt man Kapselung (*encapsulation*).

Bemerkung:

- Sowohl öffentlicher als auch privater Teil können sowohl Methoden als auch Daten enthalten.
- Öffentliche Mitglieder einer Klasse können von jeder Funktion eines Programmes benutzt werden (etwa die Methode `abheben` in `Konto`).
- Private Mitglieder können nur von den Methoden der Klasse selbst benutzt werden.

Beispiel:

```
Konto k1;
k1.abheben( -25 );          // OK
k1.k = 1000000;             // Fehler !, k private
```

Bemerkung: Kapselung erlaubt uns, das Prinzip der versteckten Information (*information hiding*) zu realisieren. David L. Parnas[16] [CACM, 15(12): 1059–1062, 1972] hat dieses Grundprinzip im Zusammenhang mit der modularen Programmierung so ausgedrückt:

1. One must provide the intended user with all the information needed to use the module correctly, and with nothing more.

2. One must provide the implementor with all the information needed to complete the module, and with nothing more.

[16] David Lorge Parnas, geb. 1941, kanadischer Informatiker.

Bemerkung: Insbesondere sollte eine Klasse alle Implementierungsdetails „verstecken“, die sich möglicherweise in Zukunft ändern werden. Da Änderungen der Implementierung meist Änderung der Datenmitglieder bedeutet, sind diese normalerweise nicht öffentlich!

Zitat: Brooks[17] [The Mythical Man-Month: Essays on Software Engineering, Addison-Wesley, 1975, page 102]:

> . . . but much more often, strategic breakthrough will come from redoing the representation of the data or tables. This is where the heart of a program lies.

Regel: Halte Datenstrukturen geheim!

[17] Fred Brooks, geb. 1931, amerik. Informatiker.

Bemerkung:

- Die „Geheimhaltung“ durch die Trennung **public**/**private** ist kein perfektes Verstecken der Implementation, weil der Benutzer der Klasse ja die Klassendefinition einsehen kann/muss.

- Sie erzwingt jedoch bei gutwilligen Benutzern ein regelkonformes Verwenden der Bibliothek.

- Andererseits schützt sie nicht gegenüber böswilligen Benutzern! (z. B. sollte man nicht erwarten, dass ein Benutzer der Bibliothek ein **private**-Feld `password` nicht auslesen kann!)

Konstruktoren und Destruktoren

Objekte werden – wie jede Variable – erzeugt und zerstört, sei es automatisch oder unter Programmiererkontrolle.

Diese Momente erfordern oft spezielle Beachtung, so dass $jede$ Klasse die folgenden Operationen zur Verfügung stellt:

- Mindestens einen Konstruktor, der aufgerufen wird, nachdem der Speicher für ein Objekt bereitgestellt wurde. Der Konstruktor hat die Aufgabe, die Datenmitglieder des Objektes geeignet zu initialisieren.

- Einen Destruktor, der aufgerufen wird, bevor der vom Objekt belegte Speicher freigegeben wird. Der Destruktor kann entsprechende Aufräumarbeiten durchführen (Beispiele folgen).

Bemerkung:

- Ein Konstruktor ist eine Methode mit demselben Namen wie die Klasse selbst und kann mit beliebigen Argumenten definiert werden. Er hat $keinen$ Rückgabewert.

- Ein Destruktor ist eine Methode, deren Name mit einer Tilde $\sim$ beginnt, gefolgt vom Namen der Klasse. Ein Destruktor hat weder Argumente noch einen Rückgabewert.

- Gibt der Programmierer keinen Konstruktor und/oder Destruktor an, so erzeugt der Übersetzer Default-Versionen. Der Default-Konstruktor hat keine Argumente.

Beispiel: Ein Beispiel für eine Klassendefinition mit Konstruktor und Destruktor:

```
class Konto
{
public:
  Konto( int start );      // Konstruktor
  ~Konto();                // Destruktor
  int kontostand();
  int abheben( int betrag );
private:
  int k;
};
```

Der Konstruktor erhält ein Argument, welches das Startkapital des Kontos sein soll (Implementierung folgt gleich). Erzeugt wird so ein Konto mittels

```
Konto k1( 1000 );  // Argumente des Konstruktors nach Objektname
Konto k2;  // Fehler! Klasse hat keinen argumentlosen Konstruktor
```

Implementierung der Klassenmethoden

Bisher haben wir noch nicht gezeigt, wie die Klassenmethoden implementiert werden. Dies ist Absicht, denn wir wollten deutlich machen, dass man nur die Definition einer Klasse *und die Semantik ihrer Methoden* wissen muss, um sie zu verwenden.

Nun wechseln wir auf die Seite des Implementierers einer Klasse. Hier nun ein vollständiges Programm mit Klassendefinition und Implementierung der Klasse `Konto`:

Programm: **(Konto.cc)**

```
#include "fcpp.hh"

class Konto
{
public:
  Konto( int start );      // Konstruktor
  ~Konto();                // Destruktor
  int kontostand();
  int abheben( int betrag );
private:
  int bilanz;
};

Konto::Konto( int startkapital )
{
```

```
    bilanz = startkapital;
    print( "Konto mit ", bilanz, " eingerichtet", 0 );
}

Konto::~Konto()
{
    print( "Konto mit ", bilanz, " aufgelöst", 0 );
}

int Konto::kontostand()
{
    return bilanz;
}

int Konto::abheben( int betrag )
{
    bilanz = bilanz - betrag;
```

```
  return bilanz;
}

int main()
{
  Konto k1( 100 ), k2( 200 );

  k1.abheben( 50 );
  k2.abheben( 300 );
}
```

Bemerkung:

- Die Definitionen der Klassenmethoden sind normale Funktionsdefinitionen, nur der Funktionsname lautet

  <Klassenname>::<Methodenname>

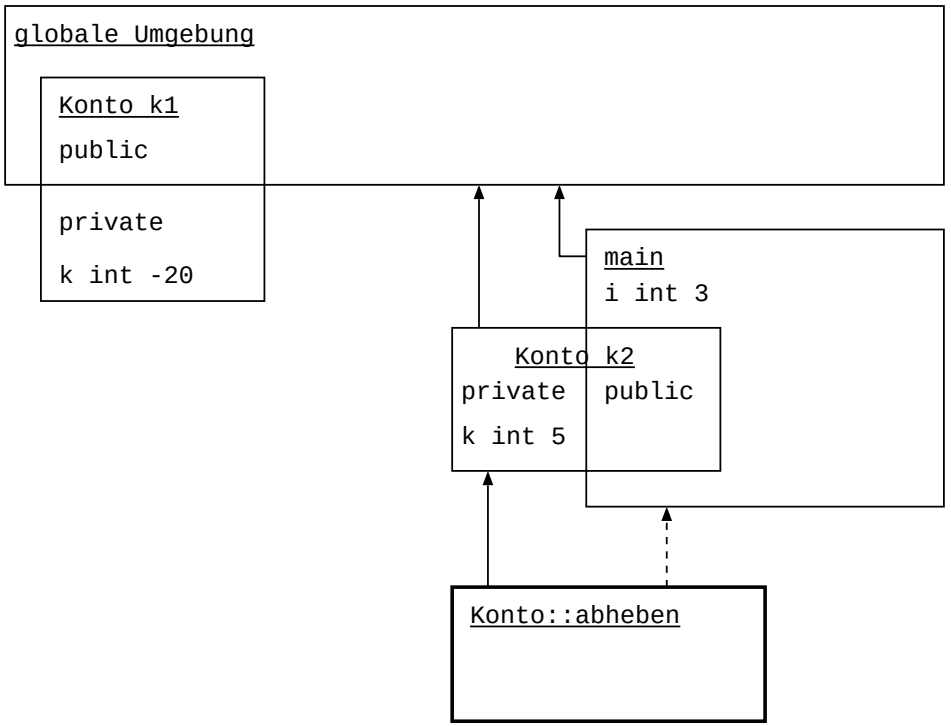
- Klassen bilden einen eigenen $Namensraum$. So ist `abheben` keine global sichtbare Funktion. Der Name `abheben` ist nur innerhalb der Definition von `Konto` sichtbar.

- Außerhalb der Klasse ist der Name erreichbar, wenn ihm der Klassenname gefolgt von zwei Doppelpunkten ($scope\ resolution\ operator$) vorangestellt wird.

Klassen im Umgebungsmodell

```
class Konto; // wie oben

Konto k1( 0 );

void main()
{
  int i = 3;
  Konto k2( 0 );

  k1.abheben( 20 );
  k2.abheben( -5 );
}
```

In `k2.abheben(-5)`



Bemerkung:

- Jedes Objekt definiert eine eigene Umgebung.
- Die öffentlichen Daten einer Objektumgebung überlappen mit der Umgebung, in der das Objekt definiert ist, und sind dort auch sichtbar.
- Der Methodenaufruf erzeugt eine neue Umgebung unterhalb der Umgebung des zugehörigen Objektes

Folgerung:

- Öffentliche Daten von `k1` sind global sichtbar.
- Öffentliche Daten von `k2` sind in `main` sichtbar.

- Private Daten von `k1` und `k2` sind von Methoden der Klasse `Konto` zugreifbar (jede Methode eines Objektes hat Zugriff auf die Mitglieder aller Objekte dieser Klasse, sofern bekannt).

Bemerkung: Die Lebensdauer von Objekten (bzw. Objektvariablen) ist genauso geregelt wie die von anderen Variablen.

Beispiel: Monte-Carlo objektorientiert

Wir betrachten nochmal das Beispiel der Bestimmung von π mit Hilfe von Zufallszahlen.

Bestandteile:

- Zufallsgenerator: Liefert bei Aufruf eine Zufallszahl.
- Experiment: Führt das Experiment einmal durch und liefert im Erfolgsfall 1, sonst 0.
- Monte-Carlo: Führt Experiment N mal durch und berechnet relative Häufigkeit.

Zufallsgenerator

Programm: Der Zufallsgenerator lässt sich hervorragend als Klasse formulieren. Er kapselt die aktuelle Zufallszahl als internen Zustand.

```
class Zufall
{
public:
  Zufall( unsigned int anfang );
  unsigned int ziehe_zahl();
private:
  int x;
};

Zufall::Zufall( unsigned int anfang )
{
  x = anfang;
```

```
}

// Implementierung ohne lange Arithmetik
// siehe Numerical Recipes, Kap. 7.
unsigned int Zufall::ziehe_zahl()
{
  // a = 7^5, m = 2^31-1
  const int ia = 16807,  im = 2147483647;
  const int iq = 127773, ir = 2836;
  const int k = x / iq;
  x = ia*(x-k*iq) - ir*k;
  if ( x < 0 ) x = x + im;
  return x;
}
```

Vorteile:

- Durch die Angabe des Konstruktors ist sichergestellt, dass der Zufallsgenerator initialisiert werden muss. Beachte: Wenn ein Konstruktor angegeben ist, so gibt es keinen Default-Konstruktor!

- Die Realisierung des Zufallsgenerators ist nach außen nicht sichtbar (`x` ist **private**). Beispielsweise könnte man nun problemlos die Implementation so abändern, dass man intern mit längeren Zahlen arbeitet.

Klasse für das Experiment

Programm:

```
class Experiment
{
public:
  Experiment( Zufall& z ); // Konstruktor
  int durchfuehren();      // einmal ausfuehren
private:
  Zufall& zg; // Merke Zufallsgenerator
  unsigned int ggT( unsigned int a, unsigned int b );
};

Experiment::Experiment( Zufall& z ) : zg( z ) {}

unsigned int Experiment::ggT( unsigned int a,
```

```
                                        unsigned int b )
{
  if ( b == 0) return a;
  else         return ggT( b, a % b );
}

int Experiment::durchfuehren()
{
  unsigned int x1 = zg.ziehe_zahl();
  unsigned int x2 = zg.ziehe_zahl();
  if ( ggT( x1, x2 ) == 1 )
    return 1;
  else
    return 0;
}
```

Bemerkung: Die Klasse `Experiment` enthält (eine Referenz auf) ein Objekt einer Klasse als Unterobjekt. Für diesen Fall gibt es eine spezielle Form des Konstruktors, die weiter unten erläutert wird.

Monte-Carlo-Funktion und Hauptprogramm

Programm:

```
#include "fcpp.hh"          // fuer print
#include "Zufall.cc"        // Code fuer die beiden
#include "Experiment.cc"    // Klassen hereinziehen

double montecarlo( Experiment& e, int N )
{
  int erfolgreich = 0;

  for ( int i=0; i<N; i=i+1 )
    erfolgreich = erfolgreich + e.durchfuehren();

  return ((double) erfolgreich) / ((double) N);
}
```

```
int main( int argc, char *argv[] )
{
  Zufall z( 93267 ); // ein Zufallsgenerator
  Experiment e( z ); // ein Experiment

  print( sqrt( 6.0/montecarlo( e,
                               readarg_int( argc,
                                            argv,
                                            1 ) ) ) );
}
```

Diskussion:

- Es gibt keine globale Variable mehr! `Zufall` kapselt den Zustand intern.
- Wir könnten auch mehrere unabhängige Zufallsgeneratoren haben.
- Die Funktion `montecarlo` kann nun mit dem Experiment parametrisiert werden. Dadurch kann man das Experiment leicht austauschen: beispielsweise erhält man π auch, indem man Punkte in $(-1, 1)^2$ würfelt und misst, wie oft sie im Einheitskreis landen.

Initialisierung von Unterobjekten

Ein Objekt kann Objekte anderer Klassen als Unterobjekte enthalten. Um in diesem Fall die ordnungsgemäße Initialisierung des Gesamtobjekts sicherzustellen, gibt es eine erweiterte Form des Konstruktors selbst.

Syntax: **(Erweiterter Konstruktor)** Ein Konstruktor für eine Klasse mit Unterobjekten hat folgende allgemeine Form:

<Konstruktor> ::= <Klassenname>::<Klassenname> (<ArgListe>) :
<UnterObjekt> (<ArgListe>)
{, <UnterObjekt> (<ArgListe>)}
{ <Rumpf> }

Die Aufrufe nach dem : sind Konstruktoraufrufe für die Unterobjekte. Deren Argumente sind Ausdrücke, die die formalen Parameter des Konstruktors des Gesamtobjektes enthalten können.

Eigenschaften:

- Bei der Ausführung jedes Konstruktors (egal ob einfacher, erweiterter oder default) werden $erst$ die Konstruktoren der Unterobjekte ausgeführt und dann der Rumpf des Konstruktors.

- Wird der Konstruktoraufruf eines Unterobjektes im erweiterten Konstruktor weggelassen, so wird dessen argumentloser Konstruktor aufgerufen. Gibt es keinen solchen, wird ein Fehler gemeldet.

- Beim Destruktor wird erst der Rumpf abgearbeitet, dann werden die Destruktoren der Unterobjekte aufgerufen. Falls man keinen programmiert hat, wird die Default-Version verwendet.

- Dies nennt man hierarchische Konstruktion/Destruktion.

Erinnerung: Eingebaute Datentypen und Zeiger haben keine Konstruktoren und werden nicht initialisiert (es sei denn man initialisiert sie explizit).

Anwendung: `Experiment` enthält eine Referenz als Unterobjekt. Mit einer Instanz der Klasse `Experiment` wird auch diese Referenz erzeugt. Referenzen müssen aber *immer* initialisiert werden, daher muss die erweiterte Form des Konstruktors benutzt werden.

Es ist in diesem Fall nicht möglich, die Referenz im Rumpf des Konstruktors zu initialisieren.

Frage: Was würde sich ändern, wenn man ein `Zufall`-Objekt statt der Referenz speichern würde?

Selbstreferenz

Innerhalb jeder Methode einer Klasse `T` ist ein Zeiger **this** vom Typ `T*` definiert, der auf das Objekt zeigt, dessen Methode aufgerufen wurde.

Beispiel: Folgendes Programmfragment zeigt eine gleichwertige Implementierung von `abheben`:

```
int Konto::abheben( int betrag )
{
  this->k = this->k - betrag;
  return this->k; // neuer Kontostand
}
```

Bemerkung: Anders ausgedrückt, ist die alte Form von `abheben` syntaktischer Zucker für die Form mit **this**. Der Nutzen von **this** wird sich später zeigen (Verkettung von Operationen).

Überladen von Funktionen und Methoden

C++ erlaubt es, mehrere Funktionen *gleichen Namens* aber mit unterschiedlicher Signatur (Zahl und Typ der Argumente) zu definieren.

Beispiel:

```
int summe()                                  { return 0; }
int summe( int i )                           { return i; }
int summe( int i, int j )                    { return i + j; }
double summe( double a, double b )           { return a + b; }

int main()
{
  int i[2];
  double x[2];
  short c;
```

```
    i[1] = summe();                  // erste Version
    i[1] = summe( 3 );               // zweite Version
    i[0] = summe( i[0], i[1] );      // dritte Version
    x[0] = summe( x[0], x[1] );      // vierte Version
    i[0] = summe( i[0], c );         // dritte Version
    i[0] = summe( x[0], i[1] );      // Fehler, mehrdeutig
}
```

Dabei bestimmt der Übersetzer anhand der Zahl und Typen der Argumente, welche Funktion aufgerufen wird. Der Rückgabewert ist dabei unerheblich.

Bezeichnung: Diesen Mechanismus nennt man Überladen von Funktionen.

Automatische Konversion

Schwierigkeiten entstehen durch automatische Konversion eingebauter numerischer Typen. Der Übersetzer geht nämlich in folgenden Stufen vor:

1. Versuche passende Funktion ohne Konversion oder mit trivialen Konversionen (z. B. Feldname nach Zeiger) zu finden. Man spricht von exakter Übereinstimmung. Dies sind die ersten vier Versionen oben.

2. Versuche innerhalb einer Familie von Typen ohne Informationsverlust zu konvertieren und so eine passende Funktion zu finden. Z. B. ist erlaubt, **bool** nach **int**, **short** nach **int**, **int** nach **long**, **float** nach **double**, etc. Im obigen Beispiel wird c in Version 5 nach **int** konvertiert.

3. Versuche Standardkonversionen (Informationsverlust!) anzuwenden: **int** nach **double**, **double** nach **int** usw.

4. Gibt es verschiedene Möglichkeiten auf $einer$ der vorigen Stufen, so wird ein Fehler gemeldet.

Tip: Verwende Überladen möglichst nur so, dass die Argumente mit einer der definierten Signaturen exakt übereinstimmen!

Überladen von Methoden

Auch Methoden einer Klasse können überladen werden. Dies benutzt man gerne für den Konstruktor, um mehrere Möglichkeiten der Initialisierung eines Objektes zu ermöglichen:

```
class Konto
{
public:
  Konto();                  // Konstruktor 1
  Konto( int start );       // Konstruktor 2
  int konto_stand();
  int abheben( int betrag );
private:
  int k;                    // Zustand
};
```

```
Konto::Konto() { k = 0; }
Konto::Konto( int start ) { k = start; }
```

Jetzt können wir ein Konto auf zwei Arten erzeugen:

```
Konto k1;        // Hat Wert 0
Konto k2(100); // Hundert Euro
```

Bemerkung:

- Eine Klasse muss einen Konstruktor ohne Argumente haben, wenn man Felder dieses Typs erzeugen will.
- Ein Default-Konstruktor wird nur erzeugt, wenn kein Konstruktor explizit programmiert wird.

Das Überladen von Funktionen ist eine Form von Polymorphismus womit man meint:

Eine Schnittstelle, viele Methoden.

Aber: Es ist sehr verwirrend, wenn überladene Funktionen sehr verschiedene Bedeutung haben. Dies sollte man vermeiden.

Objektorientierte und funktionale Programmierung

Folgendes Scheme-Programm ließ sich nur schlecht in C++ übertragen, weil die Erzeugung lokaler Funktionen nicht möglich war:

Programm: (lokal erzeugte Funktion in Scheme)

```
(define (inkrementierer n)
  (lambda (x)
    (+ x n)))

(map (inkrementierer 5)
     '(1 2 3))  => (6 7 8)
```

Mit den jetzt verfügbaren Klassen kann diese Funktionalität dagegen nachgebildet werden:

Programm: **(Inkrementierer.cc)**

```
#include "fcpp.hh"          // fuer print

class Inkrementierer
{
public:
  Inkrementierer( int n ) { inkrement = n; }
  int eval( int n ) { return n + inkrement; }
private:
  int inkrement;
};

void schleife( Inkrementierer& ink )
{
  for ( int i=1; i<10; i++ )
    print( ink.eval( i ) );
```

```
}

int main()
{
  Inkrementierer ink( 10 );
  schleife( ink );
}
```

Bemerkung:

- Man beachte die Definition der Methoden innerhalb der Klasse. Dies ist zwar kürzer, legt aber die Implementation der Schnittstelle offen.

- Die innerhalb einer Klasse definierten Methoden werden „inline“ (d. h. ohne Funktionsaufruf) übersetzt. Bei Änderungen solcher Methoden muss daher aufrufender Code neu übersetzt werden!

- Man sollte dieses Feature daher nur mit Vorsicht verwenden (z. B. bei nur lokal verwendeten Klassen oder wenn das Inlining gewünscht wird).

- Eine erweiterte Schnittstelle zur Simulation funktionaler Programme erhält man in der STL (*Standard Template Library*) mit **#include** <functional>.

Operatoren

In C++ hat man auch bei selbstgeschriebenen Klassen die Möglichkeit, einem Ausdruck wie `a+b` eine Bedeutung zu geben:

Idee: Interpretiere den Ausdruck `a+b` als `a.operator+(b)`, d. h. die Methode `operator+` des Objektes `a` (des linken Operanden) wird mit dem Argument `b` (rechter Operand) aufgerufen:

```
class X
{
public:
  X operator+( X b );
};

X X::operator+( X b ) { .... }
X a, b, c;
c = a + b;
```

Bemerkung:

- `operator+` ist also ein ganz normaler Methodenname, nur die Methode wird aus der Infixschreibweise heraus aufgerufen.

- Diese Technik ist insbesondere bei Klassen sinnvoll, die mathematische Konzepte realisieren, wie etwa rationale Zahlen, Vektoren, Polynome, Matrizen, gemischtzahlige Artihmetik, Arithmetik beliebiger Genauigkeit.

- Man sollte diese Technik zurückhaltend verwenden. Zum Beispiel sollte man + nur überladen, wenn die Operation wirklich eine Addition im mathematischen Sinn ist.

- Auch eckige Klammern `[]`, Dereferenzierung `->`, Vergleichsoperatoren `<`, `>`, `==` und sogar die Zuweisung `=` können (um-)definiert werden. `<<`, `>>` spielt bei Ein-/Ausgabe eine Rolle.

Anwendung: rationale Zahlen objektorientiert

Definition der Klasse (`Rational.hh`):

```
class Rational
{
private:
  int n, d;
  int ggT( int a, int b );
public:
  // (lesender) Zugriff auf Zaehler und Nenner
  int numerator();
  int denominator();

  // Konstruktoren
  Rational( int num, int denom ); // rational
  Rational( int num );            // ganz
```

```
    Rational();                          // Null

    // Ausgabe
    void print();

    // Operatoren
    Rational operator+( Rational q );
    Rational operator-( Rational q );
    Rational operator*( Rational q );
    Rational operator/( Rational q );
};
```

Programm: Implementierung der Methoden (`Rational.cc`):

```
int Rational::numerator()
{
  return n;
}

int Rational::denominator()
{
  return d;
}

void Rational::print()
{
  ::print( n, "/", d, 0 );
}
```

```
// ggT zum kuerzen
int Rational::ggT( int a, int b )
{
  return ( b == 0 ) ? a : ggT( b, a % b );
}

// Konstruktoren
Rational::Rational( int num, int denom )
{
  int t = ggT( num, denom );
  if ( t != 0 )
  {
    n = num / t;
    d = denom / t;
  }
  else
  {
```

```
    n = num;
    d = denom;
  }
}

Rational::Rational( int num )
{
  n = num;
  d = 1;
}

Rational::Rational()
{
  n = 0;
  d = 1;
}
```

```
// Operatoren
Rational Rational::operator+( Rational q )
{
  return Rational( n*q.d + q.n*d, d*q.d );
}

Rational Rational::operator-( Rational q )
{
  return Rational( n*q.d - q.n*d, d*q.d );
}

Rational Rational::operator*( Rational q )
{
  return Rational( n*q.n, d*q.d );
}

Rational Rational::operator/( Rational q )
```

```
{
  return Rational( n*q.d, d*q.n );
}
```

Programm: Lauffähiges Beispiel (`UseRational.cc`):

```
#include "fcpp.hh"          // fuer print
#include "Rational.hh"
#include "Rational.cc"

int main()
{
  Rational p( 3, 4 ), q( 5, 3 ), r;

  p.print(); q.print();
  r = ( p + q*p ) * p*p;
  r.print();

  return 0;
}
```

Bemerkung:

- Es ist eine gute Idee die Definition der Klasse (Schnittstelle) und die Implementierung der Methoden in getrennte Dateien zu schreiben. Dafür haben sich in C++ die Dateiendungen .hh („Headerdatei“) und .cc („Quelldatei“) eingebürgert. (Auch: .hpp, .hxx, .h, .cpp, .cxx).

- Später wird dies die sog. „getrennte Übersetzung“ ermöglichen.

- Wie schon früher erwähnt, ist die Implementierung einer leistungsfähigen gemischtzahligen Arithmetik eine hochkomplexe Aufgabe, für welche die Klasse `Rational` nur ein erster Ansatz sein kann.

- Sehr notwendig wäre auf jeden Fall die Verwendung von Ganzzahlen beliebiger Länge anstatt von **int** als Bausteine für `Rational`.

Wiederholung

Monte-Carlo Beispiel objektorientiert → Klassen helfen globale Variablen zu vermeiden.

Hierarchische Konstruktion/Destruktion von Unterobjekten. Wichtig z. B. bei Initialisierung von Referenzen.

Selbstreferenz → jede Klasse hat einen **this** Zeiger. Ist später wichtig bei Vererbung.

Überladen von Funktionen und Methoden. Nützlich z. B. bei Konstruktoren oder gemischter Arithmetik.

Operatoren erlauben infix Schreibweise.

Klassen erlauben die Übergabe von Funktionen als Parameter (später Funktoren).

Trennung von Schnittstelle und Implementierung in separaten Dateien.

Beispiel: Turingmaschine

Ein großer Vorteil der objektorientierten Programmierung ist, dass man seine Programme sehr „problemnah“ formulieren kann. Als Beispiel zeigen wir, wie man eine Turingmaschine realisieren könnte. Diese besteht aus den drei Komponenten

- Band
- Programm
- eigentliche Turingmaschine

Es bietet sich daher an, diese Einheiten als Klassen zu definieren.

Band

Programm: **(Band.hh)**

```
// Klasse fuer ein linksseitig begrenztes Band
// einer Turingmaschine.
// Das Band wird durch eine Zeichenkette aus
// Elementen des Typs char realisiert
class Band
{
public:
  // Initialisiere Band mit s, fuelle Rest
  // mit dem Zeichen init auf.
  // Setze aktuelle Bandposition auf linkes Ende.
  Band( std::string s, char init );

  // Lese Symbol unter dem Lesekopf
  char lese();
```

```
  // Schreibe und gehe links
  void schreibe_links( char symbol );

  // Schreibe und gehe rechts
  void schreibe_rechts( char symbol );

  // Drucke aktuellen Bandinhalt bis zur
  // maximal benutzten Position
  void drucke();
private:
  enum { N = 100000 };   // maximal nutzbare Groesse
  char band[N];          // das Band
  int pos;               // aktuelle Position
  int benutzt;           // bisher beschriebener Teil
};
```

TM-Programm

Programm: **(Programm.hh)**

```
// Eine Klasse, die das Programm einer
// Turingmaschine realisiert.
// Zustaende sind vom Typ int
// Bandalphabet ist der Typ char
// Anfangszustand ist Zustand in der ersten Zeile
// Endzustand ist Zustand in der letzten Zeile
class Programm
{
public:
  // Symbole fuer links/rechts
  enum R { links, rechts };

  // Erzeuge leeres Programm
```

```
Programm();

// definiere Zustandsuebergaenge
// Mit Angabe des Endzustandes ist die
// Programmierphase beendet
void zeile( int q_ein, char s_ein,
            char s_aus, R richt, int q_aus );
void zeile( int endzustand );

// lese Zustandsuebergang in Abhaengigkeit
// von akt. Zustand und gelesenem Symbol
char Ausgabe( int zustand, char symbol );
R Richtung( int zustand, char symbol );
int Folgezustand( int zustand, char symbol );

// Welcher Zustand ist Anfangszustand
int Anfangszustand();
```

```
  // Welcher Zustand ist Endzustand
  int Endzustand();

private:
  // Finde die Zeile zu geg. Zustand/Symbol
  // Liefere true, falls so eine Zeile gefunden
  // wird, sonst false
  bool FindeZeile( int zustand, char symbol );

  enum { N = 1000 };     // maximale Anzahl Uebergaenge
  int zeilen;            // Anzahl Zeilen in Tabelle
  bool fertig;           // Programmierphase beendet
  int Qaktuell[N];       // Eingabezustand
  char eingabe[N];       // Eingabesymbol
  char ausgabe[N];       // Ausgabesymbol
  R richtung[N];         // Ausgaberichtung
```

```
    int Qfolge[N];          // Folgezustand
    int letztesQ;           // Merke Zustand, Eingabe
    char letzteEingabe;     // und Zeilennummer des
    int letzteZeile;        // letzten Zugriffes.
};
```

Bemerkung: Man beachte die Definition des lokalen Datentyps `R` durch `enum`. Andererseits wird eine Form von `enum`, bei der den Konstanten gleich Zahlen zugewiesen werden, verwendet, um die Konstante `N` innerhalb der Klasse `Programm` zur Verfügung zu stellen.

Turingmaschine

Programm: **(TM.hh)**

```
// Klasse, die eine Turingmaschine realisiert
class TM
{
public:
  // Konstruiere Maschine mit Programm
  // und Band
  TM( Programm& p, Band& b );

  // Mache einen Schritt
  void Schritt();

  // Liefere true falls sich Maschine im
  // Endzustand befindet
```

```
    bool Endzustand();

  private:
    Programm& prog;  // Merke Programm
    Band& band;      // Merke Band
    int q;           // Merke akt. Zustand
};
```

Programm: (TM.cc)

```
// Konstruiere die TM mit Programm und Band
TM::TM( Programm& p, Band& b ) : prog( p ), band( b )
{
  q = p.Anfangszustand();
}

// einen Schritt machen
```

```
void TM::Schritt()
{
  // lese Bandsymbol
  char s = band.lese();

  // schreibe Band
  if ( prog.Richtung( q, s ) == Programm::links )
    band.schreibe_links( prog.Ausgabe( q, s ) );
  else
    band.schreibe_rechts( prog.Ausgabe( q, s ) );

  // bestimme Folgezustand
  q = prog.Folgezustand( q, s );
}

// Ist Endzustand erreicht?
bool TM::Endzustand()
```

```
{
  if ( q == prog.Endzustand() )
    return true;
  else
    return false;
}
```

Turingmaschinen-Hauptprogramm

Programm: **(Turingmaschine.cc)**

```
#include "fcpp.hh" // fuer print

#include "Band.hh"   // Inkludiere Quelldateien
#include "Band.cc"
#include "Programm.hh"
#include "Programm.cc"
#include "TM.hh"
#include "TM.cc"

int main( int argc, char *argv[] )
{
  // Initialisiere ein Band
  Band b( "1111", '0' );
```

```
b.drucke();

// Initialisiere ein Programm
Programm p;
p.zeile( 1, '1', 'X', Programm::rechts, 2 );
p.zeile( 2, '1', '1', Programm::rechts, 2 );
p.zeile( 2, '0', 'Y', Programm::links, 3 );
p.zeile( 3, '1', '1', Programm::links, 3 );
p.zeile( 3, 'X', '1', Programm::rechts, 4 );
p.zeile( 4, 'Y', '1', Programm::rechts, 8 );
p.zeile( 4, '1', 'X', Programm::rechts, 5 );
p.zeile( 5, '1', '1', Programm::rechts, 5 );
p.zeile( 5, 'Y', 'Y', Programm::rechts, 6 );
p.zeile( 6, '1', '1', Programm::rechts, 6 );
p.zeile( 6, '0', '1', Programm::links, 7 );
p.zeile( 7, '1', '1', Programm::links, 7 );
p.zeile( 7, 'Y', 'Y', Programm::links, 3 );
```

```
    p.zeile( 8 );

    // Baue eine Turingmaschine
    TM tm( p, b );

    // Simuliere Turingmaschine
    while ( !tm.Endzustand() )
    { // Solange nicht Endzustand
      tm.Schritt() ;              // mache einen Schritt
      b.drucke();                 // und drucke Band
    }

    return 0;                        // fertig.
}
```

Die TM realisiert das Programm „Verdoppeln einer Einserkette" von Seite 72.

Experiment: Ausgabe des oben angegebenen Programms:

```
4  Symbole auf Band initialisiert
[1]111
Programm mit  14  Zeilen definiert
Anfangszustand  1
Endzustand  8
X[1]11
X1[1]1
X11[1]
X111[0]
X11[1]Y
X1[1]1Y
X[1]11Y
[X]111Y
1[1]11Y
1X[1]1Y
1X1[1]Y
```

```
1X11[Y]
1X11Y[O]
1X11[Y]1
1X1[1]Y1
1X[1]1Y1
1[X]11Y1
11[1]1Y1
11X[1]Y1
11X1[Y]1
11X1Y[1]
11X1Y1[O]
11X1Y[1]1
11X1[Y]11
11X[1]Y11
11[X]1Y11
111[1]Y11
111X[Y]11
```

```
111XY[1]1
111XY1[1]
111XY11[O]
111XY1[1]1
111XY[1]11
111X[Y]111
111[X]Y111
1111[Y]111
11111[1]11
```

Kritik:

- Das Band könnte seine Größe dynamisch verändern.
- Statt eines einseitig unendlichen Bandes könnten wir auch ein zweiseitig unendliches Band realisieren.
- Das Finden einer Tabellenzeile könnte durch bessere Datenstrukturen beschleunigt werden.
- Bei Fehlerzuständen bricht das Programm nicht ab. Fehlerbehandlung ist keine triviale Sache.

Aber: Diese Änderungen betreffen jeweils nur die Implementierung einer einzelnen Klasse (`Band` oder `Programm`) und beeinflussen die Implementierung anderer Klassen nicht!

Abstrakter Datentyp

Eng verknüpft mit dem Begriff der Schnittstelle ist das Konzept des abstrakten Datentyps (ADT). Ein ADT besteht aus

- einer Menge von Objekten, und
- einem Satz von Operationen auf dieser Menge, sowie
- einer genauen Beschreibung der Semantik der Operationen.

Bemerkung:

- Das Konzept des ADT ist unabhängig von einer Programmiersprache, die Beschreibung kann in natürlicher (oder mathematischer) Sprache abgefasst werden.

- Der ADT beschreibt, was die Operationen tun, aber nicht, wie sie das tun. Die Realisierung ist also nicht Teil des ADT!

- Die Klasse ist der Mechanismus zur Konstruktion von abstrakten Datentypen in C++. Allerdings fehlt dort die Beschreibung der Semantik der Operationen! Diese kann man als Kommentar über die Methoden schreiben.

- In manchen Sprachen (z. B. Eiffel, PLT Scheme) ist es möglich, die Semantik teilweise zu berücksichtigen (Design by Contract: zur Funktionsdefinition kann man Vorbedingungen und Nachbedingungen angeben).

Beispiel 1: Positive m-Bit-Zahlen im Computer

Der ADT „Positive m-Bit-Zahl“ besteht aus

- Der Teilmenge $P_m = \{0, 1, \ldots, 2^m - 1\}$ der natürlichen Zahlen.
- Der Operation $+_m$ so dass für $a, b \in P_m$: $a +_m b = (a + b) \bmod 2^m$.
- Der Operation $-_m$ so dass für $a, b \in P_m$: $a -_m b = ((a - b) + 2^m) \bmod 2^m$.
- Der Operation $*_m$ so dass für $a, b \in P_m$: $a *_m b = (a * b) \bmod 2^m$.
- Der Operation $/_m$ so dass für $a, b \in P_m$: $a/_m b = q$, q die größte Zahl in P_m so dass $q *_m b \leq a$.

Bemerkung:

- Die Definition dieses ADT stützt sich auf die Mathematik (natürliche Zahlen und Operationen darauf).

- In C++ (auf einer 32-Bit Maschine) entsprechen **unsigned char**, **unsigned short**, **unsigned int** den Werten $m = 8, 16, 32$.

Beispiel 2: ADT Stack

- Ein Stack S über X besteht aus einer geordneten Folge von n Elementen aus X: $S = \{s_1, s_2, \ldots, s_n\}$, $s_i \in X$. Die Menge aller Stacks $\mathcal{S}$ besteht aus $allen$ $möglichen$ $Folgen$ der $Länge$ $n \geq 0$.

- Operation $new : \emptyset \to \mathcal{S}$, die einen leeren Stack erzeugt.

- Operation $empty : \mathcal{S} \to \{w, f\}$, die prüft ob der Stack leer ist.

- Operation $push : \mathcal{S} \times X \to \mathcal{S}$ zum Einfügen von Elementen.

- Operation $pop : \mathcal{S} \to \mathcal{S}$ zum Entfernen von Elementen.

- Operation $top : \mathcal{S} \to X$ zum Lesen des obersten Elementes.

- Die Operationen erfüllen folgende Regeln:

1. $empty(new()) = w$
2. $empty(push(S, x)) = f$
3. $top(push(S, x)) = x$
4. $pop(push(S, x)) = S$

Bemerkung:

- Die einzige Möglichkeit einen Stack zu erzeugen ist die Operation new.
- Die Regeln erlauben uns formal zu zeigen, welches Element nach einer beliebigen Folge von $push$ und pop Operationen zuoberst im Stack ist:

$$top(pop(push(push(push(new(), x_1), x_2), x_3))) =$$

$$top(push(push(new(), x_1), x_2)) = x_2$$

- Auch nicht gültige Folgen lassen sich erkennen:

$$pop(pop(push(new(), x_1))) = pop(new())$$

und dafür gibt es keine Regel!

Bemerkung: Abstrakte Datentypen, wie Stack, die Elemente einer Menge X aufnehmen, heißen auch Container. Wir werden noch eine Reihe von Containern kennenlernen: Feld, Liste (in Varianten), Queue, usw.

Beispiel 3: Das Feld

Wie beim Stack wird das Feld über einer Grundmenge X erklärt. Auch das Feld ist ein Container.

Das charakteristische an einem Feld ist der indizierte Zugriff. Wir können das Feld daher als eine Abbildung einer Indexmenge $I \subset \mathbb{N}$ in die Grundmenge X auffassen.

Die $Indexmenge$ $I \subseteq \mathbb{N}$ sei beliebig, aber im folgenden fest gewählt. Zur Abfrage der Indexmenge gebe es folgende Operationen:

- Operation min liefert kleinsten Index in I.
- Operation max liefert größten Index in I.
- Operation $isMember : \mathbb{N} \to \{w, f\}$. $isMember(i)$ liefert wahr falls $i \in I$, ansonsten falsch.

Den ADT Feld definieren wir folgendermaßen:

- Ein Feld f ist eine Abbildung der Indexmenge I in die Menge der möglichen Werte X, d. h. $f : I \to X$. Die Menge aller Felder $\mathcal{F}$ ist die Menge aller solcher Abbildungen.

- Operation $new : X \to \mathcal{F}$. $new(x)$ erzeugt neues Feld mit Indexmenge I (und initialisiert mit x, siehe unten).

- Operation $read : \mathcal{F} \times I \to X$ zum Auswerten der Abbildung.

- Operation $write : \mathcal{F} \times I \times X \to \mathcal{F}$ zum Manipulieren der Abbildung.

- Die Operationen erfüllen folgende Regeln:

  1. $read(new(x), i) = x$ für alle $i \in I$.
  2. $read(write(f, i, x), i) = x$.

3. $read(write(f, i, x), j) = read(f, j)$ für $i \neq j$.

Bemerkung:

- In unserer Definition darf $I \subset \mathbb{N}$ beliebig aber fest gewählt werden. Es sind also auch nichtzusammenhängende Indexmengen erlaubt.

- Als Variante könnte man die Manipulation der Indexmenge erlauben (die Indexmenge sollte dann als weiterer ADT definiert werden).

Klassen und dynamische Speicherverwaltung

Erinnerung: Nachteile von eingebauten Feldern in C/C++:

- Ein eingebautes Feld kennt seine Größe nicht, diese muss immer extra mitgeführt werden, was ein Konsistenzproblem mit sich bringt.
- Bei dynamischen Feldern ist der Programmierer für die Freigabe des Speicherplatzes verantwortlich.
- Eingebaute Felder sind äquivalent zu Zeigern und können daher nur *by reference* übergeben werden.
- Eingebaute Felder prüfen nicht, ob der Index im erlaubten Bereich liegt.
- Manchmal bräuchte man Verallgemeinerungen, z. B. andere Indexmengen.

Klassendefinition

Unsere Feldklasse soll Elemente des Grundtyps **float** aufnehmen. Hier ist die Klassendefinition:

Programm: **(SimpleFloatArray.hh)**

```
class SimpleFloatArray
{
public:
  // Neues Feld mit s Elementen, I=[0,s-1]
  SimpleFloatArray( int s, float f );

  // Copy-Konstruktor
  SimpleFloatArray( const SimpleFloatArray& );

  // Zuweisung von Feldern
  SimpleFloatArray& operator=(const SimpleFloatArray&);
```

```
// Destruktor: Gebe Speicher frei
~SimpleFloatArray();

// Indizierter Zugriff auf Feldelemente
// keine Ueberpruefung ob Index erlaubt
float& operator[]( int i );

// Anzahl der Indizes in der Indexmenge
int numIndices();

// kleinster Index
int minIndex();

// größter Index
int maxIndex();
```

```
    // Ist der Index in der Indexmenge?
    bool isMember( int i );

  private:
    int n;          // Anzahl Elemente
    float* p;       // Zeiger auf built-in array
};
```

Bemerkung: Man beachte, dass diese Implementierung das eingebaute Feld nutzt.

Konstruktor

Programm: **(SimpleFloatArrayImp.cc)**

```
SimpleFloatArray::SimpleFloatArray( int s, float v )
{
  n = s;
  try
  {
    p = new float[n];
  }
  catch ( std::bad_alloc )
  {
    n = 0;
    throw;
  }
  for ( int i=0; i<n; i=i+1 ) p[i] = v;
```

```
}

SimpleFloatArray::~SimpleFloatArray() { delete[] p; }

int SimpleFloatArray::numIndices() { return n; }

int SimpleFloatArray::minIndex() { return 0; }

int SimpleFloatArray::maxIndex() { return n - 1; }

bool SimpleFloatArray::isMember( int i )
{
  return ( i>=0 && i<n );
}
```

Ausnahmen

Bemerkung:

- Oben kann in der Operation **new** das Ereignis eintreten, dass nicht genug Speicher vorhanden ist. Dann setzt diese Operation eine sogenannte Ausnahme (*exception*), die in der **catch**-Anweisung abgefangen wird.

- Gute Fehlerbehandlung (Reaktionen auf Ausnahmen) ist in einem großen, professionellen Programm sehr wichtig!

- Die Schwierigkeit ist, dass man oft an der Stelle des Erkennens des Ereignisses nicht weiss, wie man darauf reagieren soll.

- Hier wird in dem Objekt vermerkt, dass das Feld die Größe 0 hat und auf eine Fehlerbehandlung an anderer Stelle verwiesen.

Indizierter Zugriff

Erinnerung: Die Operationen $read$ und $write$ des ADT Feld werden bei eingebauten Feldern durch den Operator `[]` und die Zuweisung realisiert:

```
x = 3 * a[i] + 17.5;
a[i] = 3 * x + 17.5;
```

Unsere neue Klasse soll sich in dieser Beziehung wie ein eingebautes Feld verhalten. Dies gelingt durch die Definition eines Operators `operator[]`:

Programm: **(SimpleFloatArrayIndex.cc)**

```
float& SimpleFloatArray::operator[](  int  i  )
{
  return p[i];
}
```

Bemerkung:

- `a[i]` bedeutet, dass der `operator[]` von `a` mit dem Argument `i` aufgerufen wird.

- Der Rückgabewert von `operator[]` muss eine Referenz sein, damit `a[i]` auf der linken Seite der Zuweisung stehen kann. Wir wollen ja das i-te Element des Feldes verändern und keine Kopie davon.

Copy–Konstruktor

Schließlich ist zu klären, was beim Kopieren von Feldern passieren soll. Hier sind zwei Situationen zu unterscheiden:

1. Es wird ein *neues* Objekt erzeugt welches mit einem existierenden Objekt initialisiert wird. Dies ist der Fall bei
   - Funktionsaufruf mit call by value: der aktuelle Parameter wird auf den formalen Parameter kopiert.
   - Objekt wird als Funktionswert zurückgegeben: Ein Objekt wird in eine temporäre Variable im Stack des Aufrufers kopiert.
   - Initialisierung von Objekten mit existierenden Objekten bei der Definition, also `SimpleFloatArray a(b); SimpleFloatArray a = b;`

2. Kopieren eines Objektes *auf ein bereits existierendes Objekt*, das ist die Zuweisung.

Im ersten Fall wird von C++ der sogenannte Copy-Konstruktor aufgerufen. Ein Copy-Konstruktor ist ein Konstruktor der Gestalt

<Klassenname> (const <Klassenname> &);

Als Argument wird also eine Referenz auf ein Objekt desselben Typs übergeben. Dabei bedeutet **const**, dass das Argumentobjekt nicht manipuliert werden darf.

Programm: (SimpleFloatArrayCopyCons.cc)

```
SimpleFloatArray::SimpleFloatArray( const SimpleFloatArray& a )
{
  n = a.n;
  p = new float[n];
  for ( int i=0; i<n; i=i+1 )
    p[i] = a.p[i];
}
```

Bemerkung:

- Unser Copy-Konstruktor allokiert ein neues Feld und kopiert alle Elemente des Argumentfeldes.

- Damit gibt es *immer nur jeweils einen Zeiger auf ein dynamisch erzeugtes, eingebautes Feld*. Der Destruktor kann dieses eingebaute Feld gefahrlos löschen!

Beispiel:

```
int f()
{
  SimpleFloatArray a( 100, 0.0 ); // Feld mit 100 Elem.
  SimpleFloatArray b = a;         // Aufruf Copy-Konstr.
  ... // mach etwas schlaues
} // Destruktoren rufen delete[] ihres eingeb. Feldes auf
```

Bemerkung:

- Hier hat man mit der dynamischen Speicherverwaltung der eingebauten Felder nichts mehr zu tun, und es können auch keine Fehler passieren.

- Dieses Verhalten des Copy-Konstruktors nennt man *deep copy*.

- Alternativ könnte der Copy-Konstruktor nur den Zeiger in das neue Objekt kopieren (*shallow copy*). Hier dürfte der Destruktor das Feld aber nicht einfach freigeben, weil noch Referenzen bestehen könnten! (Abhilfen: *reference counting*, *garbage collection*)

Zuweisungsoperator

Bei einer Zuweisung `a = b` soll das Objekt rechts des =-Zeichens auf das *bereits initialisierte* Objekt links des =-Zeichens kopiert werden. In diesem Fall ruft C++ den `operator=` des links stehenden Objektes mit dem rechts stehenden Objekt als Argument auf.

Programm: (SimpleFloatArrayAssign.cc)

```
SimpleFloatArray& SimpleFloatArray::operator=
  ( const SimpleFloatArray& a )
{
  // nur bei verschiedenen Objekten ist was tun
  if ( &a != this )
  {
    if ( n != a.n )
    {
      // allokiere fuer this ein
```

```
        // Feld der Groesse a.n
        delete[] p; // altes Feld loeschen
        n = a.n;
        p = new float[n]; // keine Fehlerbeh.
      }
      for ( int i=0; i<n; i=i+1 ) p[i] = a.p[i];
    }

    // Gebe Referenz zurueck damit a = b = c klappt
    return *this;
  }
```

Bemerkung:

- Haben beide Felder unterschiedliche Größe, so wird für das Feld links vom Zuweisungszeichen ein neues eingebautes Feld der korrekten Größe erzeugt.

- Der Zuweisungsoperator ist in C/C++ so definiert, dass er gleichzeitig den zugewiesenen Wert hat. Somit werden Ausdrücke wie `a = b = 0` oder `return tabelle[i] = n` möglich.

Hauptprogramm

Programm: **(UseSimpleFloatArray.cc)**

```
#include <iostream>
#include "SimpleFloatArray.hh"
#include "SimpleFloatArrayImp.cc"
#include "SimpleFloatArrayIndex.cc"
#include "SimpleFloatArrayCopyCons.cc"
#include "SimpleFloatArrayAssign.cc"

void show( SimpleFloatArray f )
{
  std::cout << "#( ";
  for ( int i=f.minIndex(); i<=f.maxIndex(); i++ )
    std::cout << f[i] << " ";
  std::cout << ")" << std::endl;
```

```
}

int main()
{
  SimpleFloatArray a( 10, 0.0 ); // erzeuge Felder
  SimpleFloatArray b( 5, 5.0 );

  for ( int i=a.minIndex(); i<=a.maxIndex(); i++ )
    a[i] = i;

  show( a ); // call by value, ruft Copy-Konstruktor
  b = a;     // ruft operator= von b
  show( b );

  // hier wird der Destruktor beider Objekte gerufen
}
```

Bemerkung:

- Jeder Aufruf der Funktion `show` kopiert das Argument mittels des Copy-Konstruktors. (Für Demonstrationszwecke: eigentlich sollte man in `show` eine Referenz verwenden!)

- Entscheidend ist, dass der Benutzer gar nicht mehr mit dynamischer Speicherverwaltung konfrontiert wird.

- Hier wird erstmals „richtige“ C++ Ausgabe verwendet. Das werden wir noch behandeln.

Default-Methoden

Für folgende Methoden einer Klasse `T` erzeugt der Übersetzer automatisch Default-Methoden, sofern man keine eigenen definiert:

- Argumentloser Konstruktor `T();`
  Dieser wird erzeugt, wenn man keinen anderen Konstruktor außer dem Copy-Konstruktor angibt. Hierarchische Konstruktion von Unterobjekten.

- Copy-Konstruktor `T( const T& );`
  Kopiert alle Mitglieder in das neue Objekt (*memberwise copy*) unter Benutzung von deren Copy-Konstruktor.

- Destruktor `~T();` Hierarchische Destruktion von Unterobjekten.

- Zuweisungsoperator `T& operator=( const T& );`
  Kopiert alle Mitglieder des Quellobjektes auf das Zielobjekt unter Nutzung der jeweiligen Zuweisungsoperatoren.

- Adress-of-Operator (`&`) mit Standardbedeutung.

Bemerkung:

- Der Konstruktor (ob default oder selbstdefiniert) ruft rekursiv die Konstruktoren von selbstdefinierten Unterobjekten auf.

- Ebenso der Destruktor.

- Enthält ein Objekt Zeiger auf andere Objekte und ist für deren Speicherverwaltung verantwortlich, so wird man wahrscheinlich alle oben genannten Methoden speziell schreiben müssen (außer dem &-Operator). Die Klasse `SimpleFloatArray` illustriert dies.

C++ Ein- und Ausgabe

Eingabe von Daten in ein Programm sowie deren Ausgabe ist ein elementarer Aspekt von Programmen.

Wir haben diesen Aspekt bis jetzt verschoben nicht weil er unwichtig wäre, sondern weil sich die entsprechenden Konstrukte in C++ nur im Kontext von Klassen und Operatoren verstehen lassen.

Jedoch werden wir hier nur die Ein- und Ausgabe von Zeichen, insbesondere auf eine Konsole, betrachten. Dies lässt sich leicht auf Dateien erweitern.

Graphische Ein- und Ausgabe werden wir aus Zeitgründen nicht betrachten. Allerdings wurde die Programmierung von graphischen Benutzerschnittstellen durch objektorientierte Programmierung revolutioniert und C++ ist dafür gut geeignet.

Namensbereiche

Zuvor benötigen wir noch das für große Programme wichtige Konstrukt der Namensbereiche welches auch in der Standardbibliothek verwendet wird.

In der globalen Umgebung darf jeder Name höchstens einmal vorkommen. Dabei ist egal, ob es sich um Namen für Variablen, Klassen oder Funktionen handelt.

Damit ergibt sich insbesondere ein Problem, wenn zwei Bibliotheken die gleichen Namen verwenden.

Eine Bibliothek ist eine Sammlung von Klassen und/oder Funktionen, die einem bestimmten Zweck dienen und von einem Programmierer zur Verfügung gestellt werden. Eine Bibliothek enthält keine main-Funktion!

Mittels Namensbereichen lässt sich dieses Problem lösen.

```
namespace A
{
  int n = 1;
  int m = n;
}

namespace B
{
  int n = 2;
  class X {};
}

int main()
{
  A::n = B::n + 3;
  return A::n;
}
```

Mittels **namespace** werden ähnlich einem Block Unterumgebungen innerhalb der globalen Umgebung geschaffen. Allerdings existieren diese Umgebungen gleichzeitig und sind auch nur innerhalb der globalen Umgebung möglich.

Namen innerhalb eines Namensbereiches werden von ausserhalb durch Voranstellen des Namens des Namensraumes und dem :: angesprochen (Qualifizierung).

Innerhalb des Namensraumes ist keine Qualifizierung erforderlich. Mittels **using** kann man sich die Qualifizierung innerhalb eines Blockes sparen.

Namensräume können wieder Namensräume enthalten.

Elemente eines Namensraumes können an verschiedenen Stellen, sogar in verschiedenen Dateien definiert werden. Ein Name darf aber innerhalb eines Namensraumes nur einmal vorkommen.

```
namespace C
{
  double x;
  int f( double x ) { return x; } // eine Funktion
  namespace D
  {
    double x, y; // x verdeckt das x in C
  }
}

namespace C
{ // fuege weitere Namen hinzu
  double y;
}

int main()
{
```

```
    C::x = 0.0; C::y = 1.0; C::D::y = 2.0;
    C::f( 2.0 );
    return 0;
}
```

Ein- und Ausgabe mit Streams

Für die Ein- und Ausgabe stellt C++ eine Reihe von Klassen und globale Variablen in der Standardbibliothek zur Verfügung. Ein- und Ausgabe ist also kein Teil der Sprache C++ selbst.

Alle Variablen, Funktionen und Klassen der C++-Standardbibliothek sind innerhalb des Namensraumes `std` definiert.

Grundlegend für die Ein- und Ausgabe in C++ ist die Idee eines Datenstromes. Dabei unterscheidet man Eingabe- und Ausgabeströme:

- Ein Ausgabestrom ist ein Objekt in welches man Datenelemente hineinsteckt. Diese werden dann an den gewünschten Ort weitergeleitet, etwa den Bildschirm oder eine Datei.

- Ein Eingabestrom ist ein Objekt aus welchem man Datenelemente herausholen kann. Diese kommen von einem gewünschten Ort, etwa der Tastatur oder einer

Datei.

Datenströme werden mittels Klassen realisiert:

- `std::istream` realisiert Eingabeströme.
- `std::ostream` realisiert Ausgabeströme.

Um diese zu verwenden, muss der Header `iostream` eingebunden werden.

Die Ein-/Ausgabe wird mittels überladener Methoden realisiert:

- **operator**>> für die Eingabe.
- **operator**<< für die Ausgabe.

Zur Ein- und Ausgabe auf der Konsole sind globale Variablen vordefiniert:

- `std::cin` vom Typ `std::istream` für die Eingabe.
- `std::cout` vom Typ `std::ostream` für die reguläre Ausgabe eines Programmes.
- `std::cerr` vom Typ `std::ostream` für die Ausgabe von Fehlern.

Damit sind wir bereit für ein Beispiel.

Programm: **(iostreamexample.cc)**

```
#include <iostream>

int main()
{
  int n;
  std::cin >> n;   // d.h. cin.operator>>(n);
  double x;
  std::cin >> x;   // d.h. cin.operator>>(x);
  std::cout << n; // d.h. cout.operator<<(n);
  std::cout << " ";
  std::cout << x;
  std::cout << std::endl; // neue Zeile
  std::cout << n << " " << x << std::endl;
  return 0;
}
```

Die Ausgabe mehrerer Objekte innerhalb einer Anweisung gelingt dadurch, dass die Methode **operator**<< den Stream, also sich selbst, als Ergebnis zurückliefert:

```
std::cout << n << std::endl;
```

ist dasselbe wie

```
( std::cout.operator<<( n ) ).operator<<( std::endl );
```

Die Methoden **operator**>> und **operator**<< sind für alle eingebauten Datentypen wie **int** oder **double** überladen.

Durch Überladen der Funktion

std :: ostream& **operator**<<(std::ostream&, <Typ>);

kann man obige Form der Ausgabe für selbstgeschriebene Klassen ermöglichen.

Als Beispiel betrachten wir eine Ausgabefunktion für die Klasse Rational:

Programm: (RationalOutput.cc)

```
std::ostream& operator<<( std::ostream& s, Rational q )
{
  s << q.numerator() << "/" << q.denominator();
  return s;
}
```

Beachte, dass das Streamargument zurückgegeben wird, um die Hintereinanderausführung zu ermöglichen.

In einer „richtigen“ Version würde man die rationale Zahl als **const** Rational& q übergeben um die Kopie zu sparen. Dies würde allerdings erfordern, die Methoden numerator und denominator als **const** zu deklarieren.

Schließlich können wir damit schreiben:

Programm: **(UseRationalOutput.cc)**

```
#include <iostream>
#include "fcpp.hh"          // fuer print
#include "Rational.hh"
#include "Rational.cc"
#include "RationalOutput.cc"

int main()
{
  Rational p( 3, 4 ), q( 5, 3 );

  std::cout << p << " " << q << std::endl;
  std::cout << (p + q*p) * p*p << std::endl;
  return 0;
}
```

Vererbung

Motivation: Polynome

Definition: Ein Polynom $p : \mathbb{R} \to \mathbb{R}$ ist eine Funktion der Form

$$p(x) = \sum_{i=0}^{n} p_i x^i,$$

Wir betrachten hier nur den Fall reellwertiger Koeffizienten $p_i \in \mathbb{R}$ und verlangen $p_n \neq 0$. n heißt dann Grad des Polynoms.

Operationen:

- Konstruktion.
- Manipulation der Koeffizienten.
- Auswerten des Polynoms an einer Stelle x.

- Addition zweier Polynome

$$p(x) = \sum_{i=0}^{n} p_i x^i, \quad q(x) = \sum_{j=0}^{m} q_j x^j$$

$$r(x) = p(x) + q(x) = \sum_{i=0}^{\max(n,m)} \underbrace{(p_i^* + q_i^*)}_{r_i} x^i$$

$$p_i^* = \begin{cases} p_i & i \leq n \\ 0 & \text{sonst} \end{cases}, \quad q_i^* = \begin{cases} q_i & i \leq m \\ 0 & \text{sonst} \end{cases}.$$

- Multiplikation zweier Polynome

$$
\begin{aligned}
r(x) = p(x) * q(x) &= \left(\sum_{i=0}^{n} p_i x^i\right)\left(\sum_{j=0}^{m} q_j x^j\right) \\
&= \sum_{i=0}^{n}\sum_{j=0}^{m} p_i q_j x^{i+j} \\
&= \sum_{k=0}^{m+n} \underbrace{\left(\sum_{\{(i,j)|i+j=k\}} p_i q_j\right)}_{r_k} x^k
\end{aligned}
$$

Implementation

In großen Programmen möchte man Codeduplizierung vermeiden und möglichst viel Code wiederverwenden.

Für den Koeffizientenvektor $p_0, \ldots, p_n$ wäre offensichtlich ein Feld der adäquate Datentyp. Wir wollen unser Feld `SimpleFloatArray` benutzen. Eine Möglichkeit:

Programm:

```
class Polynomial
{
private:
  SimpleFloatArray coefficients;
public:
    ...
};
```

Alternativ kann man Polynome als Exemplare von `SimpleFloatArray` mit zusätzlichen Eigenschaften ansehen, was im folgenden ausgeführt wird.

Öffentliche Vererbung

Programm: Definition der Klasse `Polynomial` mittels Vererbung:
(Polynomial.hh)

```
class Polynomial : public SimpleFloatArray
{
public:
  // konstruiere Polynom vom Grad n
  Polynomial( int n );

  // Default-Destruktor ist ok
  // Default-Copy-Konstruktor ist ok
  // Default-Zuweisung ist ok

  // Grad des Polynoms
  int degree();
```

```
    // Auswertung
    float eval( float x );

    // Addition von Polynomen
    Polynomial operator+( Polynomial q );

    // Multiplikation von Polynomen
    Polynomial operator*( Polynomial q );

    // Gleichheit
    bool operator==( Polynomial q );

    // drucke Polynom
    void print();
};
```

Syntax: Die Syntax der öffentlichen Vererbung lautet:

<ÖAbleitung> ::= class <Klasse2> : public <Klasse1> {
<Rumpf>
} ;

Bemerkung:

- Klasse 1 heißt Basisklasse, Klasse 2 heißt abgeleitete Klasse.
- Klasse 2 enthält ein Objekt von Klasse 1 als Unterobjekt.
- Alle öffentlichen Mitglieder der Basisklasse mit Ausnahme von Konstruktoren, Destruktor und Zuweisungsoperatoren sind auch öffentliche Mitglieder der abgeleiteten Klasse. Sie operieren auf dem Unterobjekt.
- Im Rumpf kann Klasse 2 weitere Mitglieder vereinbaren.

- Daher spricht man auch von einer Erweiterung einer Klasse durch öffentliche Ableitung.

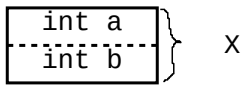
- Alle privaten Mitglieder der Basisklasse sind *keine* Mitglieder der Klasse 2. Damit haben auch Methoden der abgeleiteten Klasse *keinen* Zugriff auf private Mitglieder der Basisklasse.

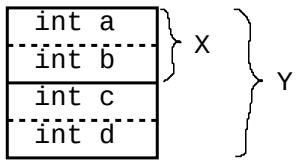
- Eine Klasse kann mehrere Basisklassen haben (Mehrfachvererbung), diesen Fall behandeln wir hier aber nicht.

Beispiel zu public/private und öffentlicher Vererbung

```
class X
{
public:
  int a;
  void A();
private:
  int b;
  void B();
};
```

```
class Y : public X
{
public:
  int c;
  void C();
private:
  int d;
  void D();
};
```





```
X x;

x.a = 5;      // OK
x.b = 10;     // Fehler

void X::A()
{
  B();        // OK
  b = 3;      // OK
}
```

```
Y y;
y.a = 1;  // OK
y.c = 2;  // OK
y.b = 4;  // Fehler
y.d = 8;  // Fehler

void Y::C() {
  d = 8;  // OK
  b = 4;  // Fehler
  A();    // OK
  B();    // Fehler
}
```

Ist-ein-Beziehung

Ein Objekt einer abgeleiteten Klasse enthält ein Objekt der Basisklasse als Unterobjekt.

Daher darf ein Objekt der abgeleiteten Klasse für ein Objekt der Basisklasse eingesetzt werden. Allerdings sind dann nur Methoden der Basisklasse für das Objekt aufrufbar.

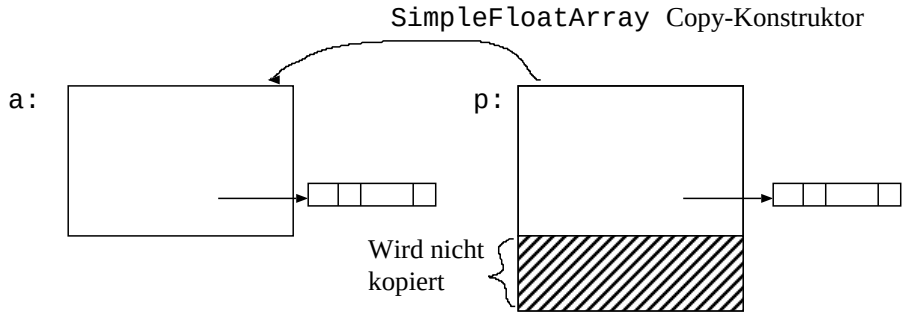
Beispiel:

```
void g( SimpleFloatArray a ) { a[3] = 1.0; }

Polynomial p( 10 );
SimpleFloatArray b( 100, 0.0 );
g( p );   // (1) OK
p = b;    // (2) Fehler
b = p;    // (3) OK
```

Bemerkung:

- Im Fall (1) wird bei Aufruf von `g(p)` der Copy-Konstruktor des formalen Parameters `a`, also `SimpleFloatArray`, benutzt, um das `SimpleFloatArray`-Unterobjekt von `p` auf den formalen Parameter `a` vom Typ `SimpleFloatArray` zu kopieren.

- Falls `Polynomial` weitere Datenmitglieder hätte, so würde die Situation so aussehen:



In diesem Fall spricht man von *slicing*.

- Im Fall (2) soll einem Objekt der abgeleiteten Klasse ein Objekt der Basisklasse zugewiesen werden. Dies ist nicht erlaubt, da nicht klar ist, welchen Wert etwaige zusätzliche Datenmitglieder der abgeleiteten Klasse bekommen sollen.

- Fall (3) ist OK, der Zuweisungsoperator der Basisklasse wird aufgerufen und das Unterobjekt aus der abgeleiteten Klasse dem links stehenden Objekt der Basisklasse zugewiesen.

Konstruktoren, Destruktor und Zuweisungsoperatoren

Programm: **(PolynomialKons.cc)**

```
Polynomial::Polynomial( int n ) :
  SimpleFloatArray( n+1, 0.0 ) {}
```

Bemerkung:

- Die syntaktische Form entspricht der Initialisierung von Unterobjekten wie oben beschrieben.

- Die Implementierung des Copy-Konstruktors kann man sich sparen, da der Default-Copy-Konstruktor das Gewünschte leistet, dasselbe gilt für Zuweisungsoperator und Destruktor.

Auswertung

Programm: Auswertung mit Horner-Schema **(PolynomialEval.cc)**

```
// Auswertung
float Polynomial::eval( float x )
{
  float sum = 0.0;

  // Hornerschema
  for ( int i=maxIndex(); i>=0; i=i-1 )
    sum = sum * x + operator[](i);
  return sum;
}
```

Bemerkung: Statt **operator**[](i) könnte man (∗**this**)[i] schreiben.

Weitere Methoden

Programm: (PolynomialImp.cc)

```
// Grad auswerten
int Polynomial::degree()
{
  return maxIndex();
}

// Addition von Polynomen
Polynomial Polynomial::operator+( Polynomial q )
{
  int nr = degree(); // mein Grad

  if ( q.degree() > nr) nr = q.degree();

  Polynomial r( nr ); // Ergebnispolynom
```

```
  for ( int i=0; i<=nr; i=i+1 )
  {
    if ( i <= degree() )
      r[i] = r[i] + (*this)[i]; // add me to r
    if ( i <= q.degree() )
      r[i] = r[i] + q[i];       // add q to r
  }

  return r;
}

// Multiplikation von Polynomen
Polynomial Polynomial::operator*( Polynomial q )
{
  Polynomial r( degree() + q.degree() ); // Ergebnispolynom

  for ( int i=0; i<=degree(); i=i+1 )
    for ( int j=0; j<=q.degree(); j=j+1 )
      r[i+j] = r[i+j] + (*this)[i] * q[j];
```

```
  return r;
}

// Drucken
void Polynomial::print()
{
  if ( degree() < 0 )
    std::cout << 0;
  else
    std::cout << (*this)[0];

  for ( int i=1; i<=maxIndex(); i=i+1 )
    std::cout << "+" << (*this)[i] << "*x^" << i;

  std::cout << std::endl;
}
```

Gleichheit

Gleichheit ist kein einfaches Konzept, wie man ja schon an Zahlen sieht: ist `0 == 0.0`? Oder `(int) 1000000000 == (short) 1000000000`? Gleichheit für selbstdefinierte Datentypen ist daher Sache des Programmierers:

Programm: (PolynomialEqual.cc)

```
bool Polynomial::operator==( Polynomial q )
{
  if ( q.degree() > degree() )
  {
    for ( int i=0; i<=degree(); i=i+1 )
      if ( (*this)[i] != q[i] ) return false;
    for ( int i=degree()+1; i<=q.degree(); i=i+1 )
      if ( q[i] != 0.0 ) return false;
  }
  else
```

```
  {
    for ( int i=0; i<=q.degree(); i=i+1 )
      if ( (*this)[i] != q[i] ) return false;
    for ( int i=q.degree()+1; i<=degree(); i=i+1 )
      if ( (*this)[i] != 0.0 ) return false;
  }

  return true;
}
```

Bemerkung: Im Gegensatz dazu ist Gleichheit von Zeigern immer definiert. Zwei Zeiger sind gleich, wenn sie auf *dasselbe* Objekt zeigen:

```
Polynomial p( 10 ), q( 10 );

Polynomial* z1 = &p;
Polynomial* z2 = &p;
Polynomial* z3 = &q;

if ( z1 == z2 ) ... // ist wahr
if ( z1 == z3 ) ... // ist falsch
```

Benutzung von `Polynomial`

Folgendes Beispiel definiert das Polynom

$$p = 1 + x$$

und druckt p, p^2 und p^3.

Programm: **(UsePolynomial.cc)**

```
#include <iostream>

// alles zum SimpleFloatArray
#include "SimpleFloatArray.hh"
#include "SimpleFloatArrayImp.cc"
#include "SimpleFloatArrayIndex.cc"
#include "SimpleFloatArrayCopyCons.cc"
```

```
#include "SimpleFloatArrayAssign.cc"

// Das Polynom
#include "Polynomial.hh"
#include "PolynomialImp.cc"
#include "PolynomialKons.cc"
#include "PolynomialEqual.cc"
#include "PolynomialEval.cc"

int main()
{
  Polynomial p( 2 ), q( 10 );

  p[0] = 1.0;
  p[1] = 1.0;
  p.print();
```

```
    q = p * p;
    q.print();

    q = p * p * p;
    q.print();
  }
```

mit der Ausgabe:

```
1+1*x^1
1+2*x^1+1*x^2
1+3*x^1+3*x^2+1*x^3
```

Diskussion

- Diese Implementation hat die wesentliche Schwachstelle, dass der Grad bei führenden Nullen mathematisch nicht korrekt ist. Angenehmer wäre, wenn Konstanten den Grad 0 hätten, lineare Polynome den Grad 1 und der Koeffizientenvektor allgemein die Länge Grad+1 hätte.

- Die Abhilfe könnte darin bestehen, dafür zu sorgen, dass der Konstruktor nur Polynome mit korrektem Grad erzeugt. Allerdings können Polynome ja beliebig modifiziert werden, daher wäre eine andere Möglichkeit in der Methode `degree` den Grad jeweils aus den Koeffizienten zu bestimmen.

Wiederholung: Vererbung

Öffentliche Ableitung einer Klasse:

```
class B : public A { ... };
```

Öffentliche Mitglieder von A sind auch öffentliche Mitglieder von B. Private Mitglieder von A sind in B nicht sichtbar.

Objekt der Klasse B enthält Objekt der Klasse A als Unterobjekt.

Ein Objekt der Klasse B kann für ein Objekt der Klasse A eingesetzt werden, etwa bei einem Funktionsaufruf (Ist-ein Beziehung).

Konstruktor, Destruktor und Zuweisungsoperator werden nicht vererbt, es werden aber ggf. Default-Varianten erzeugt.

Private Vererbung

Wenn man nur die Implementierung von `SimpleFloatArray` nutzen will, ohne die Methoden öffentlich zu machen, so kann man dies durch private Vererbung erreichen:

Programm:

```
class Polynomial : private SimpleFloatArray
{
public:
   ...
};
```

Bemerkung: Hier müsste man dann die Schnittstelle zum Zugriff auf die Koeffizienten des Polynoms selbst Programmieren, oder zumindest die Initialisierung mittels

```
Polynomial::Polynomial( SimpleFloatArray& coeffs ) { ... }
```

Eigenschaften der privaten Vererbung

Bemerkung: Private Vererbung bedeutet:

- Ein Objekt der abgeleiteten Klasse enthält ein Objekt der Basisklasse als Unterobjekt.
- Alle *öffentlichen* Mitglieder der Basisklasse werden *private* Mitglieder der abgeleiteten Klasse.
- Alle privaten Mitglieder der Basisklasse sind keine Mitglieder der abgeleiteten Klasse.
- Ein Objekt der abgeleiteten Klasse kann *nicht* für ein Objekt der Basisklasse eingesetzt werden!

Zusammenfassung

Wir haben somit drei verschiedene Möglichkeiten kennengelernt, um die Klasse `SimpleFloatArray` für `Polynomial` zu nutzen:

1. Als privates Datenmitglied

2. Mittels öffentlicher Vererbung

3. Mittels privater Vererbung

Bemerkung: Je nach Situation ist die eine oder andere Variante angemessener. Hier hängt viel vom guten Geschmack des Programmierers ab. In diesem speziellen Fall würde ich persönlich Möglichkeit 1 bevorzugen.

Methodenauswahl und virtuelle Funktionen

Motivation: Feld mit Bereichsprüfung

Problem: Die für die Klasse `SimpleFloatArray` implementierte Methode `operator[]` prüft nicht, ob der Index im erlaubten Bereich liegt. Zumindest in der Entwicklungsphase eines Programmes wäre es aber nützlich, ein Feld mit Indexüberprüfung zu haben.

Abhilfe: Ableitung einer Klasse `CheckedSimpleFloatArray`, bei der sich `operator[]` anders verhält.

Programm: Klassendefinition **(CheckedSimpleFloatArray.hh)**:

```
class CheckedSimpleFloatArray :
  public SimpleFloatArray
{
public:
  CheckedSimpleFloatArray( int s, float f );

  // Default-Versionen von copy Konstruktor,
    Zuweisungsoperator
  // und Destruktor sind OK

  // Indizierter Zugriff mit Indexprüfung
  float& operator[]( int i );
};
```

Methodendefinition **(CheckedSimpleFloatArrayImp.cc)**:

```
CheckedSimpleFloatArray::
  CheckedSimpleFloatArray( int s, float f ) :
  SimpleFloatArray( s, f )
{}

float& CheckedSimpleFloatArray::operator[]( int i )
{
  assert( i>=minIndex() && i<=maxIndex() );
  return SimpleFloatArray::operator[](i);
}
```

Verwendung **(UseCheckedSimpleFloatArray.cc)**:

```
#include <iostream>
#include <cassert>

#include "SimpleFloatArray.hh"
#include "SimpleFloatArrayImp.cc"
#include "SimpleFloatArrayIndex.cc"
#include "SimpleFloatArrayCopyCons.cc"
#include "SimpleFloatArrayAssign.cc"

#include "CheckedSimpleFloatArray.hh"
#include "CheckedSimpleFloatArrayImp.cc"

void g( SimpleFloatArray& a )
{
  std::cout << "beginn in g: " << std::endl;
  std::cout << "access: " << a[1] << "  " << a[10] << std::endl;
  std::cout << "ende in g: " << std::endl;
```

```
}

int main()
{
  CheckedSimpleFloatArray a( 10, 0 );
  g( a );
  std::cout << "beginn in main: " << std::endl;
  std::cout << "zugriff in main: " << a[10] << std::endl;
  std::cout << "ende in main: " << std::endl;
}
```

mit der Ausgabe:

```
beginn in g:
access: 0  1.85018e-40
ende in g:
beginn in main:
UseCheckedSimpleFloatArray: CheckedSimpleFloatArrayImp.cc:8:
  float& CheckedSimpleFloatArray::operator[](int):
  Assertion `i>=minIndex() && i<=maxIndex()' failed.
```

```
Aborted (core dumped)
```

Bemerkung:

- In der Funktion `main` funktioniert die Bereichsprüfung dann wie erwartet.
- In der Funktion g wird hingegen keine Bereichsprüfung durchgeführt, auch wenn sie mit einem Objekt vom Typ `CheckedSimpleFloatArray` aufgerufen wird!
- Warum ist das so?
- In der Funktion g betrachtet der Compiler die übergebene Referenz als Objekt vom Typ SimpleFloatArray und dies hat keine Bereichsprüfung.
- Der Funktionsaufruf mit dem Objekt der öffentlich abgeleiteten Klasse ändert daran nichts! Dies ist eine Konsequenz der Realisierung der Ist-ein-Beziehung.

- Die Auswahl der Methode hängt vom angegeben Typ ab und nicht vom konkreten Objekt welches übergeben wird.

- Meistens ist dies aber nicht das gewünschte Verhalten (vgl. dazu auch die späteren Beispiele).

Virtuelle Funktionen

Idee: Gib dem Compiler genügend Information, so dass er schon bei der Übersetzung von `SimpleFloatArray`-Methoden ein flexibles Verhalten von `[]` möglich macht. In C++ geschieht dies, indem man Methoden in der Basisklasse als virtuell (*virtual*) kennzeichnet.

Programm:

```
class SimpleFloatArray
{
public:
  ...
  virtual float& operator[]( int i );
  ...
private:
  ...
};
```

Beobachtung: Mit dieser Änderung funktioniert die Bereichsprüfung auch in der Funktion `g` in `UseCheckedSimpleFloatArray.cc`: wird sie mit einer Referenz auf ein `CheckedSimpleFloatArray`-Objekt aufgerufen, so wird der Bereichstest durchgeführt, bei Aufruf mit einer Referenz auf ein `SimpleFloatArray`-Objekt aber nicht.

Bemerkung:

- Die Einführung einer virtuellen Funktion erfordert also Änderungen in bereits existierendem Code, nämlich der Definition der Basisklasse!

- Die Implementierung der Methoden bleibt jedoch unverändert.

Implementation: Diese Auswahl der Methode in Abhängigkeit vom tatsächlichen Typ des Objekts kann man dadurch erreichen, dass jedes Objekt entweder Typinformation oder einen Zeiger auf eine Tabelle mit den für seine Klasse virtuell definierten Funktionen mitführt.

Bemerkung:

- Wird eine als virtuell markierte Methode in einer abgeleiteten Klasse neu implementiert, so wird die Methode der abgeleiteten Klasse verwendet, wenn das Objekt für ein Basisklassenobjekt eingesetzt wird.

- Die Definition der Methode in der abgeleiteten Klasse muss genau mit der Definition in der Basisklasse übereinstimmen, ansonsten wird überladen!

- Das Schlüsselwort **virtual** muss in der abgeleiteten Klasse nicht wiederholt werden, es ist aber guter Stil dies zu tun.

- Die Eigenschaften virtueller Funktionen lassen sich nur nutzen, wenn auf das

Objekt über Referenzen oder Zeiger zugegriffen wird! Bei einem Aufruf (*call-by-value*) von

```
void g( SimpleFloatArray a )
{
  cout << a[1] << "  " << a[11] << endl;
}
```

erzeugt der Copy-Konstruktor ein Objekt `a` vom Typ `SimpleFloatArray` (Slicing!) und innerhalb von `g()` wird entsprechend dessen `operator[]` verwendet.

- Virtuelle Funktionen stellen wieder eine Form des Polymorphismus dar („eine Schnittstelle — viele Methoden“).

- Der Zugriff auf eine Methode über die Tabelle virtueller Funktionen ist deutlich ineffizienter, was für Objektorientierung auf niedriger Ebene eine Rolle spielen kann.

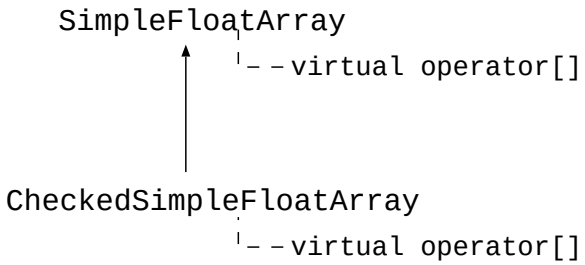
- In vielen objektorientierten Sprachen (z. B. Smalltalk, Objective C, Common Lisp/CLOS) verhalten sich alle Methoden „virtuell“.

- In der Programmiersprache Java ist das virtuelle Verhalten der Normalfall, das Default-Verhalten von C++-Methoden kann man aber durch Hinzufügen des Schlüsselworts `static` erreichen.

Abstrakte Klassen

Motivation

Hatten:



Beobachtung: Beide Klassen besitzen dieselben Methoden und unterscheiden sich nur in der Implementierung von `operator[]`. Wir könnten ebenso `SimpleFloatArray` von `CheckedSimpleFloatArray` ableiten. Das Klassendiagramm drückt diese Symmetrie aber nicht aus.

Grund: `SimpleFloatArray` stellt sowohl die Definition der Schnittstelle eines ADT *Feld* dar, als auch eine Implementierung dieser Schnittstelle. Es ist aber sinnvoll, diese beiden Aspekte zu trennen.

Schnittstellenbasisklassen

Idee: Definiere eine möglichst allgemeine Klasse `FloatArray`, von der sowohl `SimpleFloatArray` als auch `CheckedSimpleFloatArray` abgeleitet werden.

Bemerkung: Oft will und kann man für (virtuelle) Methoden in einer solchen Basisklasse keine Implementierung angeben. In C++ kennzeichnet man sie dann mit dem Zusatz `= 0` am Ende. Solche Funktionen bezeichnet man als rein virtuelle (*engl.: pure virtual*) Funktionen.

Beispiel: **(FloatArray.hh)**

```
class FloatArray
{
public:
  virtual ~FloatArray() {};
  virtual float& operator[]( int i ) = 0;
  virtual int numIndices() = 0;
  virtual int minIndex() = 0;
  virtual int maxIndex() = 0;
  virtual bool isMember( int i ) = 0;
};
```

Bezeichnung: Klassen, die mindestens eine rein virtuelle Funktion enthalten, nennt man abstrakt. Das Gegenteil ist eine konkrete Klasse.

Bemerkung:

- Man kann keine Objekte von abstrakten Klassen instanzieren. Aus diesem Grund haben abstrakte Klassen auch keine Konstruktoren.

- Sehr wohl kann man aber Zeiger und Referenzen dieses Typs haben, die dann aber auf Objekte abgeleiteter Klassen zeigen.

- Eine abstrakte Klasse, die der Definition einer Schnittstelle dient, bezeichnen wir nach Barton/Nackman als Schnittstellenbasisklasse (*interface base class*).

- Schnittstellenbasisklassen enthalten üblicherweise keine Datenmitglieder und alle Methoden sind rein virtuell.

- Die Implementierung dieser Schnittstelle erfolgt in abgeleiteten Klassen.

Bemerkung: (Virtueller Destruktor) Eine Schnittstellenbasisklasse sollte einen virtuellen Destruktor

```
virtual ~FloatArray();
```

mit einer Dummy-Implementierung

```
FloatArray::~FloatArray() {}
```

besitzen, damit man dynamisch erzeugte Objekte abgeleiter Klassen durch die Schnittstelle der Basisklasse löschen kann. Beispiel:

```
void g( FloatArray* p )
{
  delete p;
}
```

Bemerkung: Der Destruktor darf nicht rein virtuell sein, da der Destruktor abgeleiteter Klassen einen Destruktor der Basisklasse aufrufen will.

Beispiel: geometrische Formen

Aufgabe: Wir wollen mit zweidimensionalen geometrischen Formen arbeiten. Dies sind von einer Kurve umschlossene Flächen wie Kreis, Rechteck, Dreieck, ...

Programm: Eine mögliche C++-Implementierung wäre folgende **(shape.cc)**:

```
#include <iostream>
#include <cmath>

const double pi = 3.1415926536;

class Shape
{
public:
  virtual ~Shape() {};
  virtual double area() = 0;
  virtual double diameter() = 0;
  virtual double circumference() = 0;
};
```

```
// works on every shape
double circumference_to_area( Shape& shape )
{
  return shape.circumference() / shape.area();
}

class Circle : public Shape
{
public:
  Circle( double r ) { radius = r; }
  virtual double area()
  {
    return pi * radius * radius;
  }
  virtual double diameter()
  {
    return 2 * radius;
  }
```

```
  virtual double circumference()
  {
    return 2 * pi * radius;
  }
private:
  double radius;
};

class Rectangle : public Shape
{
public:
  Rectangle( double aa, double bb )
  {
    a = aa; b = bb;
  }
  virtual double area() { return a*b; }
  virtual double diameter()
  {
    return sqrt( a*a + b*b );
```

```
  }
  virtual double circumference()
  {
    return 2 * (a + b);
  }
private:
  double a, b;
};

int main()
{
  Rectangle unit_square( 1.0, 1.0 );
  Circle unit_circle( 1.0 );
  Circle unit_area_circle( 1.0 / sqrt( pi ) );

  std::cout << "Das Verhaeltnis von Umfang zu Flaeche betraegt\n";
  std::cout << "Einheitsquadrat:     "
            << circumference_to_area( unit_square )
            << std::endl;
```

```
    std::cout << "Kreis mit Flaeche 1: "
              << circumference_to_area( unit_area_circle )
              << std::endl;
    std::cout << "Einheitskreis:       "
              << circumference_to_area( unit_circle )
              << std::endl;
    return 0;
}
```

Ergebnis: Wir erhalten als Ausgabe des Programms:

```
Das Verhaeltnis von Umfang zu Flaeche betraegt
Einheitsquadrat:     4
Kreis mit Flaeche 1: 3.54491
Einheitskreis:       2
```

Beispiel: Funktoren

$\rightarrow$ Klasse der Objekte, die sich wie Funktionen verhalten.

Hatten: Definition einer Inkrementierer-Klasse in `Inkrementierer.cc`. Nachteile waren:

- Möchten beliebige Funktion auf die Listenelemente anwenden können.
- Syntax `ink.eval(...)` nicht optimal.

Dies wollen wir nun mit Hilfe einer Schnittstellenbasisklasse und der Verwendung von `operator()` verbessern.

Programm: (Funktor.cc)

```
#include <iostream>

class Function
{
public:
  virtual ~Function() {};
  virtual int operator()( int ) = 0;
};

class Inkrementierer : public Function
{
public:
  Inkrementierer( int n ) { inkrement = n; }
  virtual int operator()( int n ) { return n + inkrement; }
private:
  int inkrement;
};
```

```
void schleife( Function& func )
{
  for ( int i=1; i<10; i++ )
    std::cout << func( i ) << " ";
  std::cout << std::endl;
}

class Quadrat : public Function
{
public:
  virtual int operator()( int n ) { return n * n; }
};

int main()
{
  Inkrementierer ink( 10 );
  Quadrat quadrat;
  schleife( ink );
```

```
    schleife( quadrat );
}
```

Bemerkung: Unangenehm ist jetzt eigentlich nur noch, dass der Argument- und Rückgabetyp der Methode auf **int** festgelegt ist. Dies wird bald durch Schablonen (*Templates*) behoben werden.

Wiederholung: Virtuelle Funktionen

```
class A { public: virtual void f() {} };

class B : public A { public: virtual void f() {} };

void g( A& a ) { a.f(); }
```

Abgeleitete Klasse definiert Methode mit gleichem Namen und Argumenten wie in der Basisklasse.

In g darf Objekt der Klasse A oder B eingesetzt werden.

Mit **virtual** wird in der Funktion g die Methodenauswahl am konkreten Objekt vorgenommen.

Ohne **virtual** wird in der Funktion g die Methodenauswahl anhand des Typs A vorgenommen.

Achtung: Ohne Referenz wird in jedem Fall eine Kopie erstellt und die Methode der Klasse A verwendet.

Abstrakte Klassen haben mindestens eine rein virtuelle Methode.

Schnittstellenbasisklassen dienen zur Definition einer Schnittstelle und haben in der Regel nur rein virtuelle Methoden und keine Datenmitglieder.

Beispiel: Exotische Felder

Programm: Wir definieren folgende (schon gezeigte) Schnittstellenbasisklasse **(FloatArray.hh)**:

```
class FloatArray
{
public:
  virtual ~FloatArray() {};
  virtual float& operator[]( int i ) = 0;
  virtual int numIndices() = 0;
  virtual int minIndex() = 0;
  virtual int maxIndex() = 0;
  virtual bool isMember( int i ) = 0;
};
```

Von dieser kann man leicht `SimpleFloatArray` ableiten. Außerdem passt die Schnittstelle auf weitere Feldtypen, was wir im Folgenden zeigen wollen.

Dynamisches Feld

Wir wollen jetzt ein Feld mit variabel großer, aber zusammenhängender Indexmenge $I = \{o, o+1, \ldots, o+n-1\}$ mit $o, n \in \mathbb{Z}$ und $n \geq 0$ definieren. Wir gehen dazu folgendermaßen vor:

- Der Konstruktor fängt mit einem Feld der Länge 0 an ($o = n = 0$).
- `operator[]` prüft, ob $i \in I$ gilt; wenn nein, so wird der Indexbereich erweitert, ein entsprechendes Feld allokiert, die Werte aus dem alten Feld werden in das neue kopiert und das alte danach freigegeben.

Programm: (DFA.cc)

```
class DynamicFloatArray : public FloatArray
{
public:
  DynamicFloatArray() { n = 0; o = 0; p = new float[1]; }
  virtual ~DynamicFloatArray() { delete[] p; }
  virtual float& operator[]( int i );
  virtual int numIndices() { return n; }
  virtual int minIndex() { return o; }
  virtual int maxIndex() { return o + n - 1; }
  virtual bool isMember( int i ) { return (i >= o) && (i < o+n); }
private:
  int n;        // Anzahl Elemente
  int o;        // Ursprung der Indexmenge
  float* p;     // Zeiger auf built-in array
};

float& DynamicFloatArray::operator[]( int i )
```

```
{
  if ( i < o || i >= o+n )
  {     // resize
    int new_o, new_n;
    if ( i < o ) {
      new_o = i;
      new_n = n + o - i;
    }
    else
    {
      new_o = o;
      new_n = i - o + 1;
    }
    float* q = new float[new_n];
    for ( int j=0; j<new_n; j=j+1 ) q[j] = 0.0;
    for ( int j=0; j<n; j=j+1 )
      q[j + o - new_o] = p[j];
    delete[] p;
    p = q;
```

```
      n = new_n;
      o = new_o;
    }
    return p[i - o];
  }
```

Bemerkung:

- Im Konstruktor wird der Einfachheit halber bereits ein **float** allokiert.
- Der Copy-Konstruktor und Zuweisungsoperator müssten auch implementiert werden (um zu vermeiden, dass der Zeiger kopiert wird).

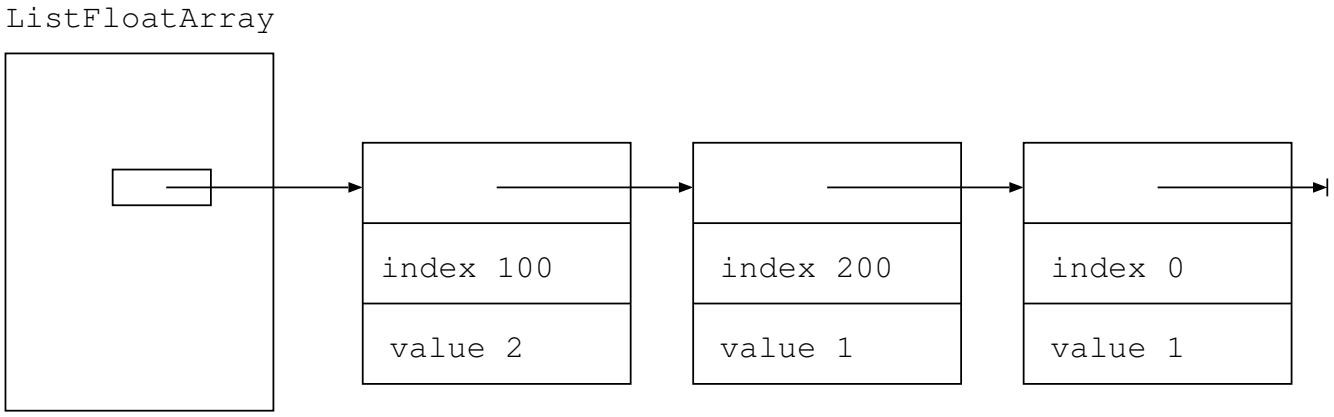
Listenbasiertes Feld

Problem: Wenn man `DynamicFloatArray` zur Darstellung von Polynomen verwendet, so werden Polynome mit vielen Nullkoeffizienten, z. B.

$$p(x) = x^{100} + 1 \text{ oder } q(x) = p^2(x) = x^{200} + 2x^{100} + 1$$

sehr ineffizient verwaltet.

Abhilfe: Speichere die Elemente des Feldes als einfach verkettete *Liste* von Index–Wert–Paaren:



Programm: (LFA.cc)

```
class ListFloatArray :
  public FloatArray
{
public:
  ListFloatArray();                    // leeres Feld
  virtual ~ListFloatArray();           // ersetzt ~FloatArray

  virtual float& operator[]( int i );
  virtual int numIndices();
  virtual int minIndex();
  virtual int maxIndex();
  virtual bool isMember( int i );
private:
  struct FloatListElem
  { // lokale Struktur
    FloatListElem* next;
    int index;
```

```
    float value;
  };
  FloatListElem* insert( int i, float v );
  FloatListElem* find( int i );

  int n;              // Anzahl Elemente
  FloatListElem* p; // Listenanfang
};

// private Hilfsfunktionen
ListFloatArray::FloatListElem*
  ListFloatArray::insert( int i, float v )
{
  FloatListElem* q = new FloatListElem;

  q->index = i;
  q->value = v;
  q->next = p;
  p = q;
```

```
  n = n + 1;
  return q;
}

ListFloatArray::FloatListElem*
  ListFloatArray::find( int i )
{
  for ( FloatListElem* q=p; q!=0; q=q->next )
    if ( q->index == i )
      return q;
  return 0;
}

// Konstruktor
ListFloatArray::ListFloatArray()
{
  n = 0; // alles leer
  p = 0;
}
```

```
// Destruktor
ListFloatArray::~ListFloatArray()
{
  while ( p != 0 )
  {
    FloatListElem* q;
    q = p;              // q ist erstes
    p = q->next;        // entferne q aus Liste
    delete q;
  }
}

float& ListFloatArray::operator[]( int i )
{
  FloatListElem* r = find( i );
  if ( r == 0 )
    r = insert( i, 0.0 ); // index einfuegen
  return r->value;
```

```
}

int ListFloatArray::numIndices()
{
  return n;
}

int ListFloatArray::minIndex()
{
  if ( p == 0 ) return 0;
  int min = p->index;
  for ( FloatListElem* q=p->next; q!=0; q=q->next )
    if ( q->index < min ) min = q->index;
  return min;
}

int ListFloatArray::maxIndex()
{
  if ( p == 0 ) return 0;
```

```
  int max = p->index;
  for ( FloatListElem* q=p->next; q!=0; q=q->next )
    if ( q->index > max ) max = q->index;
  return max;
}

bool ListFloatArray::isMember( int i )
{
  return ( find( i ) != 0 );
}
```

Bemerkung:

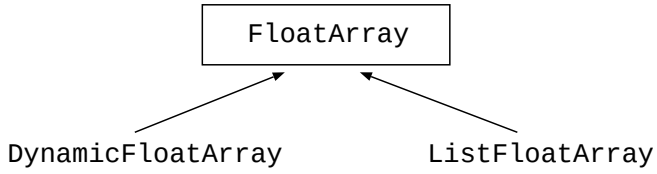
- Das Programm verwendet eine unsortierte Liste. Man hätte auch eine sortierte Liste verwenden können (Vorteile?).

- Für die Index–Wert–Paare wird innerhalb der Klassendefinition der zusammengesetzte Datentyp `FloatListElem` definiert.

- Die privaten Methoden dienen der Manipulation der Liste und werden in der Implementierung der öffentlichen Methoden verwendet. Merke: Innerhalb einer Klasse können wiederum Klassen definiert werden!

Anwendung

Wir haben somit folgendes Klassendiagramm:



Da sowohl `DynamicFloatArray` als auch `ListFloatArray` die durch `FloatArray` definierte Schnittstelle erfüllen, kann man nun Methoden für `FloatArray` schreiben, die auf beiden abgeleiteten Klassen funktionieren.

Als Beispiel betrachten wir folgendes Programm, welches `FloatArray` wieder zur Polynom-Multiplikation verwendet (der Einfachheit halber ohne es in eine Klasse `Polynomial` zu packen).

Programm: (UseFloatArray.cc)

```
#include <iostream>

#include "FloatArray.hh"
#include "DFA.cc"
#include "LFA.cc"

void polyshow( FloatArray& f )
{
  for ( int i=f.minIndex(); i<=f.maxIndex(); i=i+1 )
    if ( f.isMember( i ) && f[i] != 0.0 )
      std::cout << "+" << f[i] << "*x^" << i;
  std::cout << std::endl;
}

void polymul( FloatArray& a, FloatArray& b, FloatArray& c )
{
  // Loesche a
```

```
  for ( int i=a.minIndex(); i<=a.maxIndex(); i=i+1 )
    if ( a.isMember( i ) )
      a[i] = 0.0;

  // a = b*c
  for ( int i=b.minIndex(); i<=b.maxIndex(); i=i+1 )
    if ( b.isMember( i ) )
      for ( int j=c.minIndex(); j<=c.maxIndex(); j=j+1 )
        if ( c.isMember( j ) )
          a[i+j] = a[i+j] + b[i] * c[j];
}

int main()
{
  // funktioniert mit einer der folgenden Zeilen:
  // DynamicFloatArray f, g;
  ListFloatArray f, g;

  f[0] = 1.0; f[100] = 1.0;
```

```
    polymul( g, f, f );
    polymul( f, g, g );
    polymul( g, f, f );
    polymul( f, g, g ); // f = (1 + x^100)^16

    polyshow( f );
}
```

Ausgabe:

```
+1*x^0+16*x^1000+120*x^2000+560*x^3000+1820*x^4000+4368*x^5000+8008*x^6000
+11440*x^7000+12870*x^8000+11440*x^9000+8008*x^10000+4368*x^11000
+1820*x^12000+560*x^13000+120*x^14000+16*x^15000+1*x^16000
```

Bemerkung:

- Man kann nun sehr „spät“, nämlich erst in der `main`-Funktion entscheiden, mit welcher Art Felder man tatsächlich arbeiten will.
- Je nachdem, wie vollbesetzt der Koeffizientenvektor ist, ist entweder `DynamicFloatArray` oder `ListFloatArray` günstiger.
- Schlecht ist noch die Weise, in der allgemeine Schleifen über das Feld implementiert werden. Die Anwendung auf `ListFloatArray` ist sehr ineffektiv! Eine Abhilfe werden wir bald kennenlernen (Iteratoren).

Zusammenfassung

In diesem Abschnitt haben wir gezeigt, wie man mit Hilfe von Schnittstellenbasisklassen eine Trennung von

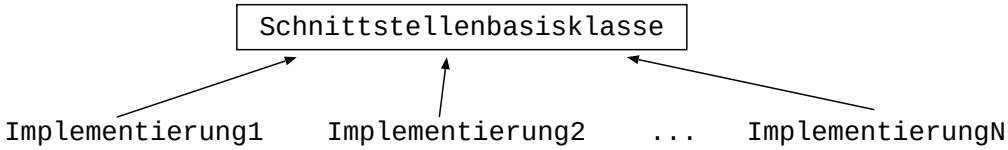
- Schnittstellendefinition und
- Implementierung

erreicht.

Dies gelingt durch

- rein virtuelle Funktionen in Verbindung mit
- Vererbung.

Typischerweise erhält man Klassendiagramme der Form:



Man *erzeugt* Objekte konkreter (abgeleiteter) Klassen und *benutzt* diese Objekte durch die Schnittstellenbasisklasse:

Create objects, use interfaces!

Generische Programmierung

Funktionsschablonen

Motivation: Auswechseln von Datentypen in streng typgebundenen Sprachen.

Definition: Eine Funktionsschablone (*Function Template*) entsteht, indem man die Präambel

```
template <class T>
```

bzw.

```
template <typename T>
```

einer Funktionsdefinition voranstellt. In der Schablonendefinition kann man `T` dann wie einen vorhandenen Datentyp verwenden. Dieser Typ muss keine Klasse sein, was die Einführung der äquivalenten Definition mittels **typename** motivierte.

Programm: Vertauschen des Inhalts zweier gleichartiger Referenzen:

```
template <class T> void swap( T& a, T& b )
{
  T t = a;
  a = b; b = t;
}

int main()
{
  int a = 10, b = 20;
  swap( a, b );
}
```

Bemerkung:

- Bei der Übersetzung von `swap(a,b)` generiert der Übersetzer die Version `swap(int& a, int& b)` und übersetzt sie (es sei denn, es gibt schon genau so

eine Funktion).

- Wie beim Überladen von Funktionen wird die Funktion nur anhand der Argumente ausgewählt. Der Rückgabewert spielt keine Rolle.

- Im Unterschied zum Überladen generiert der Übersetzer für jede vorkommende Kombination von Argumenten eine Version der Funktion (keine automatische Typkonversion).

- Dies nennt man automatische Template-Instanzierung.

Programm: Beispiel: Maximum

```
template <class T> T max( T a, T b )
{
  if ( a < b ) return b; else return a;
}
```

Bemerkung: Hier muss für den Typ `T` ein **operator**< definiert sein.

Beispiel: wieder funktionales Programmieren

Problem: Der Aufruf virtueller Funktionen erfordert Entscheidungen zur Laufzeit, was in einigen (wenigen) Fällen zu langsam sein kann.

Abhilfe: Verwendung von Funktionsschablonen.

Programm: Funktionales Programmieren mit Schablonen **(Funktional-statisch.cc)**:

```
#include <iostream>
using namespace std;

class Inkrementierer
{
public:
  Inkrementierer( int n ) { inkrement = n; }
  int operator()( int n ) { return n + inkrement; }
```

```
private:
  int inkrement;
};

class Quadrat
{
public:
  int operator()( int n ) { return n * n; }
};

template <class T> void schleife( T& func )
{
  for ( int i=1; i<10; i++ )
    cout << func(i) << " ";
  cout << endl;
}
```

```
int main()
{
  Inkrementierer ink( 10 );
  Quadrat quadrat;
  schleife( ink );
  schleife( quadrat );
}
```

Bemerkung:

- Es werden die passenden Varianten der Funktion `schleife` erzeugt.
- Unterschied zur Variante mit der gemeinsamen Basisklasse Function:
  - Statt genau einer gibt es nun mehrere Varianten der Funktion `schleife`.
  - Methodenaufrufe am Argument `func` erfolgen **nicht** über virtuelle Funktionen.
- Nachteil: Leider haben wir aber keine Schnittstellendefinition mehr.

Bezeichnung: Man nennt diese Technik auch statischen Polymorphismus, da die Methodenauswahl zur Übersetzungszeit erfolgt. Im Gegensatz dazu bezeichnet man die Verwendung virtueller Funktionen als dynamischen Polymorphismus.

Empfehlung: Wenden Sie diese oder ähnliche Techniken (wie etwa die sogenannten *Expression Templates*) nur an, wenn es unbedingt notwendig ist. Untersuchen Sie auch vorher das Laufzeitverhalten (Profiling), denn laut Donald E. Knuth (ursprünglich wohl von C. A. R. Hoare) gilt:

> Premature optimization is the root of all evil!

Klassenschablonen

Problem: Unsere selbstdefinierten Felder und Listen sind noch zu unflexibel. So hätten wir beispielsweise auch gerne Felder von **int**-Zahlen.

Bemerkung: Dieses Problem rührt von der statischen Typbindung von C/C++ her und tritt bei Sprachen mit dynamischer Typbindung (Scheme, Python, . . .) nicht auf. Allerdings ist es für solche Sprachen viel schwieriger hocheffizienten Code zu generieren.

Abhilfe: Die C++-Lösung für dieses Problem sind parametrisierte Klassen, die auch Klassenschablonen (*Class Templates*) genannt werden.

Definition: Eine Klassenschablone entsteht, indem man der Klassendefinition die Präambel **template** <**class** T> voranstellt. In der Klassendefinition kann dann der Parameter `T` wie ein Datentyp verwendet werden.

Beispiel:

```
// Schablonendefinition
template <class T> class SimpleArray
{
public:
  SimpleArray( int s, T f );
...
};
// Verwendung
SimpleArray<int> a( 10, 0 );
SimpleArray<float> b( 10, 0.0 );
```

Bemerkung:

- `SimpleArray` alleine ist kein Datentyp!
- `SimpleArray<int>` ist ein neuer Datentyp, d. h. Sie können Objekte dieses Typs

erzeugen, oder ihn als Parameter/Rückgabewert einer Funktion verwenden.

- Der Mechanismus arbeitet wieder zur Übersetzungszeit des Programmes. Bei *Übersetzung* der Zeile

```
SimpleArray<int> a( 10, 0 );
```

generiert der Übersetzer den Programmtext für `SimpleArray<int>`, der aus dem Text der Klassenschablone `SimpleArray` entsteht, indem alle Vorkommen von `T` durch **int** ersetzt werden. Anschließend wird diese Klassendefinition übersetzt.

- Da der Übersetzer selbst C++-Programmcode generiert, spricht man auch von generischer Programmierung.

- Den Vorgang der Erzeugung einer konkreten Variante einer Klasse zur Übersetzungszeit nennt man auch Template-Instanzierung. Allerdings gibt es bei Klassen, im Gegensatz zu Funktionen, keine automatische Instanzierung.

- Der Name Schablone (*Template*) kommt daher, dass man sich die parametrisierte Klasse als Schablone vorstellt, die zur Anfertigung konkreter Varianten benutzt wird.

Programm: **(SimpleArray.hh)**

```
template <class T>
class SimpleArray
{
public:
  SimpleArray( int s, T f );
  SimpleArray( const SimpleArray<T>& );
  SimpleArray<T>& operator=( const SimpleArray<T>& );
  ~SimpleArray();

  T& operator[]( int i );
  int numIndices();
```

```
  int minIndex();
  int maxIndex();
  bool isMember( int i );

private:
  int n;   // Anzahl Elemente
  T* p;    // Zeiger auf built-in array
};
```

Bemerkung: Syntaktische Besonderheiten:

- Wird die Klasse selbst als Argument oder Rückgabewert im Rumpf der Definition benötigt, schreibt man `SimpleArray<T>`.

- Im Namen des Konstruktors bzw. Destruktors taucht *kein* `T` auf. Der Klassenparameter parametrisiert den Klassennamen, nicht aber die Methodennamen.

- Die Definition des Destruktors (als Beispiel) lautet dann:

```
template <class T>
SimpleArray<T>::~SimpleArray() { delete[] p; }
```

Programm: Methodenimplementierung **(SimpleArrayImp.cc)**:

```
// Destruktor
template <class T>
SimpleArray<T>::~SimpleArray()
{
  delete[] p;
}

// Konstruktor
template <class T>
SimpleArray<T>::SimpleArray( int s, T v )
{
  n = s;
  p = new T[n];
  for ( int i=0; i<n; i=i+1 ) p[i] = v;
}
```

```
// Copy-Konstruktor
template <class T>
SimpleArray<T>::SimpleArray( const SimpleArray<T>& a )
{
  n = a.n;
  p = new T[n];
  for ( int i=0; i<n; i=i+1 )
    p[i] = a.p[i];
}

// Zuweisungsoperator
template <class T>
SimpleArray<T>& SimpleArray<T>::operator=
( const SimpleArray<T>& a )
{
  if ( &a != this )
```

```
  {
    if ( n != a.n )
    {
      delete[] p;
      n = a.n;
      p = new T[n];
    }
    for ( int i=0; i<n; i=i+1 ) p[i] = a.p[i];
  }
  return *this;
}

template <class T>
inline T& SimpleArray<T>::operator[]( int i )
{
  return p[i];
}
```

```
template <class T>
inline int SimpleArray<T>::numIndices()
{
  return n;
}

template <class T>
inline int SimpleArray<T>::minIndex()
{
  return 0;
}

template <class T>
inline int SimpleArray<T>::maxIndex()
{
  return n - 1;
```

```
}

template <class T>
inline bool SimpleArray<T>::isMember( int i )
{
  return ( i >= 0 && i < n );
}

template <class T>
std::ostream& operator<<( std::ostream& s,
                          SimpleArray<T>& a )
{
  s << "#( ";
  for ( int i=a.minIndex(); i<=a.maxIndex(); i=i+1 )
    s << a[i] << " ";
  s << ")" << std::endl;
  return s;
```

```
}
```

Programm: Verwendung **(UseSimpleArray.cc)**:

```
#include <iostream>

#include "SimpleArray.hh"
#include "SimpleArrayImp.cc"

int main()
{
  SimpleArray<float> a( 10, 0.0 ); // erzeuge
  SimpleArray<int> b( 25, 5 );     // Felder

  for ( int i=a.minIndex(); i<=a.maxIndex(); i++ )
    a[i] = i;
  for ( int i=b.minIndex(); i<=b.maxIndex(); i++ )
    b[i] = i;
```

```
    std::cout << a << std::endl << b << std::endl;

    // hier wird der Destruktor gerufen
}
```

Beispiel: Feld fester Größe

Bemerkung:

- Eine Schablone kann auch mehr als einen Parameter haben.
- Als Schablonenparameter sind nicht nur Klassennamen, sondern z. B. auch Konstanten von eingebauten Typen erlaubt.

Anwendung: Ein Feld fester Größe könnte folgendermaßen definiert und verwendet werden:

```
template <class T, int m>
class SimpleArrayCS
{
public:
  SimpleArrayCS( T f );
  ...
private:
  T p[m]; // built-in array fester Groesse
};

...

SimpleArrayCS<int, 5> a( 0 );
SimpleArrayCS<float, 3> a( 0.0 );
...
```

Bemerkung:

- Die Größe ist hier auch zur Übersetzungszeit festgelegt und muss nicht mehr gespeichert werden.
- Da nun keine Zeiger auf dynamisch allokierte Objekte verwendet werden, sind für Copy-Konstruktor, Zuweisung und Destruktor die Defaultmethoden ausreichend.
- Der Compiler kann bei bekannter Feldgröße unter Umständen effizienteren Code generieren, was vor allem für kleine Felder interessant ist (z. B. Vektoren im $\mathbb{R}^2$ oder $\mathbb{R}^3$).
- Es ist ein wichtiges Kennzeichen von C++, dass Objektorientierung bei richtigem Gebrauch auch für sehr kleine Datenstrukturen ohne Effizienzverlust angewendet werden kann.

Beispiel: Smart Pointer

Problem: Dynamisch erzeugte Objekte können ausschließlich über Zeiger verwaltet werden. Wie bereits diskutiert, ist die konsistente Verwaltung des Zeigers (bzw. der Zeiger) und des Objekts nicht einfach.

Abhilfe: Entwurf mit einem neuen Datentyp, der anstatt eines Zeigers verwendet wird. Mittels Definition von **operator**$*$ und **operator**$->$ kann man erreichen, dass sich der neue Datentyp wie ein eingebauter Zeiger benutzen lässt. In Copy-Konstruktor und Zuweisungsoperator wird dann *reference counting* eingebaut.

Bezeichnung: Ein Datentyp mit dieser Eigenschaft wird intelligenter Zeiger (*smart pointer*) genannt.

Programm: **(Zeigerklasse zum *reference counting*, Ptr.hh)**

```
template <class T>
class Ptr
{
  struct RefCntObj
  {
    int count;
    T* obj;
    RefCntObj( T* q ) { count = 1; obj = q; }
  };
  RefCntObj* p;

  void report()
  {
    std::cout << "refcnt = " << p->count << std::endl;
  }
```

```
void increment()
{
  p->count = p->count + 1;
  report();
}

void decrement()
{
  p->count = p->count - 1;
  report();
  if ( p->count == 0 )
  {
    delete p->obj; // Geht nicht fuer Felder!
    delete p;
  }
}
```

```
public:
  Ptr() { p = 0; }

  Ptr( T* q )
  {
    p = new RefCntObj( q );
    report();
  }

  Ptr( const Ptr<T>& y )
  {
    p = y.p;
    if ( p != 0 ) increment();
  }

  ~Ptr()
```

```
{
  if ( p != 0 ) decrement();
}

Ptr<T>& operator=( const Ptr<T>& y )
{
  if ( p != y.p )
  {
    if ( p != 0 ) decrement();
    p = y.p;
    if ( p != 0 ) increment();
  }
  return *this;
}

T& operator*() { return *(p->obj); }
```

```
  T* operator->() { return p->obj; }
};
```

Programm: **(Anwendungsbeispiel, PtrTest.cc)**

```
#include <iostream>
#include "Ptr.hh"

int g( Ptr<int> p )
{
  return *p;
}

int main()
{
  Ptr<int> q = new int( 17 );
  std::cout << *q << std::endl;
  int x = g( q );
  std::cout << x << std::endl;
  Ptr<int> z = new int( 22 );
```

```
    q = z;
    std::cout << *q << std::endl;
    // nun wird alles automatisch geloescht !
}
```

Bemerkung:

- Man beachte die sehr einfache Verwendung durch Ersetzen der eingebauten Zeiger (die natürlich nicht weiterverwendet werden sollten!).
- Nachteil: mehr Speicher wird benötigt (das `RefCntObj`)
- Es gibt verschiedene Möglichkeiten, *reference counting* zu implementieren, die sich bezüglich Speicher- und Rechenaufwand unterscheiden.
- Die hier vorgestellte Zeigerklasse funktioniert (wegen **delete**[]) nicht für Felder!
- *Reference counting* funktioniert nicht für Datenstrukturen mit Zykeln $\leadsto$ andere Techniken zur automatischen Speicherverwaltung notwendig.

Wiederholung: Generische Programmierung

```
template <class T>
T min( T a, T b ) { if ( a < b ) return a; else return b; }

int a, b, c;
c = min( a, b );
double x, y, z;
z = min( x, y );
```

Ermöglicht es, eine Funktion für verschiedene Datentypen zu schreiben. Übersetzer generiert eigene Version der Funktion für jeden Datentypen.

Automatische Extraktion der Template-Parameter.

Geht auch: c = min<int>(2.0, 3.0);

```
template <class T>
class SimpleArray
{
public:
  ...
  T& operator[]( int i ) { ... }
private:
  int n; T* p;
};

SimpleArray<int> a;
SimpleArray<double> x;
```

Ermöglicht es, eine Klasse mit einem Datentyp zu parametrisieren.

Übersetzer erzeugt für jeden Datentyp `T` eine eigene Version `SimpleArray<T>`. Diese sind vollkommen separate Klassen.

Template-Argument muss explizit angegeben werden.

Gemeinsame Verwendung von Klassen und Funktionsschablone:

```
template <class T>
T min( SimpleArray<T>& x )
{
  T m = x[ x.minIndex() ];
  for ( int i=x.minIndex()+1; i<=x.maxIndex(); i=i+1 )
    if ( x[i] < m ) m = x[i];
  return m;
}

SimpleArray<int> a( 10, 1 );
SimpleArray<double> x( 20, 3.14 );
int b = min( a );
```

Dies nennt man eine Spezialisierung (einer Funktionsschablone), da `min` nicht mit beliebigen Datentypen aufgerufen werden kann.

Effizienz generischer Programmierung

Beispiel: Bubblesort

Aufgabe: Ein Feld von Zahlen $a = (a_0, a_1, a_2, \ldots, a_{n-1})$ ist zu sortieren. Die Sortierfunktion liefert als Ergebnis eine Permutation $a' = (a'_0, a'_1, a'_2, \ldots, a'_{n-1})$ der Feldelemente zurück, so dass

$$a'_0 \leq a'_1 \leq \ldots \leq a'_{n-1}$$

Idee: Der Algorithmus Bubblesort ist folgendermaßen definiert:

- Gegeben sei ein Feld $a = (a_0, a_1, a_2, \ldots, a_{n-1})$ der Länge n.

- Durchlaufe die Indizes $i = 0, 1, \ldots, n-2$ und vergleiche jeweils a_i und a_{i+1}. Ist $a_i > a_{i+1}$ so vertausche die beiden. Beispiel:

	17	10	8	16
$i = 0$	10	17	8	16
$i = 1$	10	8	17	16
$i = 2$	10	8	16	17

  Am Ende eines solchen Durchlaufes steht die größte der Zahlen sicher ganz rechts und ist damit an der richtigen Position.

- Damit bleibt noch ein Feld der Länge $n - 1$ zu sortieren.

Satz: t_{cs} sei eine obere Schranke der Kosten für einen Vergleich und einen swap (Tausch), und n bezeichne die Länge des Felds. Falls t_{cs} nicht von n abhängt, so hat Bubblesort eine asymptotische Laufzeit von $O(n^2)$.

Beweis:

$$\sum_{i=0}^{n-1} \sum_{j=0}^{i-1} t_{cs} = t_{cs} \sum_{i=0}^{n-1} i = t_{cs} \frac{(n-1)n}{2} = O(n^2)$$

Programm: **(bubblesort_.cc)**

```
/* ist in namespace std schon enthalten:
   template <class T> void swap( T& a, T& b ) {
     T t = a;
     a = b;
     b = t;
   }
*/

template <class C> void bubblesort( C& a )
{
  for ( int i=a.maxIndex(); i>=a.minIndex(); i=i-1 )
    for ( int j=a.minIndex(); j<i; j=j+1 )
      if ( a[j+1] < a[j] )
        std::swap( a[j+1], a[j] );
}
```

Bemerkung:

- Die Funktion `bubblesort` benötigt, dass auf $Elementen$ des Feldes der Vergleichsoperator **operator**$<$ definiert ist.

- Die Funktion benutzt die öffentliche Schnittstelle der Feldklassen, die wir programmiert haben, d. h. für `C` können wir jede unserer Feldklassen einsetzen!

Mit folgender Routine kann man Laufzeiten verschiedener Programmteile messen:

Programm: (timestamp.cc)

```
#include <ctime>

// Setzt Marke und gibt Zeitdifferenz zur letzten
// Marke zurueck
clock_t last_time;
double time_stamp()
{
  clock_t current_time = clock();
  double duration =
    ((double) (current_time - last_time)) /
      CLOCKS_PER_SEC;
  last_time = current_time;
  return duration;
}
```

Dies wenden wir auf Bubblesort an:

Programm: Bubblesort für verschiedene Feldtypen (UseBubblesort.cc)

```
#include <iostream>

// SimpleFloatArray mit virtuellem operator[]
#include "SimpleFloatArrayV.hh"
#include "SimpleFloatArrayImp.cc"
#include "SimpleFloatArrayIndex.cc"
#include "SimpleFloatArrayCopyCons.cc"
#include "SimpleFloatArrayAssign.cc"

// templatisierte Variante mit variabler Groesse
#include "SimpleArray.hh"
#include "SimpleArrayImp.cc"

// templatisierte Variante mit Compile-Zeit Groesse
#include "SimpleArrayCS.hh"
#include "SimpleArrayCSImp.cc"
```

```
// dynamisches listenbasiertes Feld
#include "FloatArray.hh"
#include "ListFloatArrayDerived.hh"
#include "ListFloatArrayImp.cc"

// Zeitmessung
#include "timestamp.cc"

// generischer bubblesort
#include "bubblesort_.cc"

// Zufallsgenerator
#include "Zufall.cc"

const int n = 32000;
const int samples = 50; // Mittle "samples" Messungen
const int lowSamples = 10; // Anzahl fuer langsame Impl.

static Zufall z( 93576 );
```

```
template <class T>
void initialisiere( T& a )
{
  for ( int i=0; i<n; i=i+1 )
    a[i] = z.ziehe_zahl();
}

int main()
{
  SimpleArrayCS<float, n> a( 0.0 );
  SimpleArray<float> b( n, 0.0 );
  SimpleFloatArray c( n, 0.0 );
  ListFloatArray d;

  initialisiere( a ); initialisiere( b );
  initialisiere( c ); initialisiere( d );

  double duration;
```

```
duration = 0.0;
std::cout << "SimpleArrayCS ...";
time_stamp();
for ( int s=0; s<samples; s=s+1 )
{
  bubblesort( a );
  duration = duration + time_stamp();
}
std::cout << duration / samples << " sec" << std::endl;

duration = 0.0;
std::cout << "SimpleArray ...";
time_stamp();
for ( int s=0; s<samples; s=s+1 )
{
  bubblesort( b );
  duration = duration + time_stamp();
}
```

```
std::cout << duration / samples << " sec" << std::endl;

duration = 0.0;
std::cout << "SimpleFloatArray ...";
time_stamp();
for ( int s=0; s<samples; s=s+1 )
{
  bubblesort( c );
  duration = duration + time_stamp();
}
std::cout << duration / samples << " sec" << std::endl;

// duration = 0.0;
// std::cout << "ListFloatArray ...";
// time_stamp();
// for ( int s=0; s<lowSamples; s=s+1 )
// {
//   bubblesort( d );
//   duration = duration + time_stamp();
```

```
    // }
    // std::cout << duration / lowSamples << " sec" << std::endl;
}
```

Ergebnis:

Ergebnisse vom WS 2002/2003:

n	1000	2000	4000	8000	16000	32000
built-in array	0.01	0.04	0.14	0.52	2.08	8.39
`SimpleArrayCS`	0.01	0.03	0.15	0.58	2.30	9.12
`SimpleArray`	0.01	0.05	0.15	0.60	2.43	9.68
`SimpleArray` ohne inline	0.04	0.15	0.55	2.20	8.80	35.31
`SimpleFloatArrayV`	0.04	0.15	0.58	2.28	9.13	36.60
`ListFloatArray`	4.62	52.38	—	—	—	—

WS 2011/2012, gcc 4.5.0, 2.26 GHz Intel Core 2 Duo:

n	1000	2000	4000	8000	16000	32000
`SimpleArrayCS`	0.0029	0.0089	0.034	0.138	0.557	2.205
`SimpleArray`	0.0031	0.0098	0.039	0.156	0.622	2.499
`SimpleFloatArrayV`	0.0110	0.0204	0.083	0.330	1.322	5.288

WS 2014/2015, gcc 4.9.2 -O3, 2.6 GHz Intel Core i7:

n	1000	2000	4000	8000	16000	32000
`SimpleArrayCS`	0.00096	0.0032	0.013	0.070	0.328	1.4048
`SimpleArray`	0.00096	0.0034	0.013	0.0727	0.329	1.4095
`SimpleFloatArrayV`	0.0008	0.0027	0.011	0.0579	0.297	1.4353
`ListFloatArray`	1.056	8.999	—	—	—	—

WS 2015/2016, gcc 5.2.1 -O3, 4.0 GHz Intel Core i7-4790K:

n	1000	2000	4000	8000	16000	32000
`SimpleArrayCS`	0.000238	0.000936	0.00372	0.0148	0.0594	0.240
`SimpleArray`	0.000237	0.000938	0.00371	0.0148	0.0593	0.240
`SimpleFloatArrayV`	0.000185	0.000716	0.00282	0.0112	0.0447	0.182
`ListFloatArray`	0.646912	6.68187	—	—	—	—

WS 2015/2016, gcc 5.2.1, 4.0 GHz Intel Core i7-4790K:

n	1000	2000	4000	8000	16000	32000
`SimpleArrayCS`	0.00208	0.00837	0.0330	0.132	0.545	2.192
`SimpleArray`	0.00208	0.00833	0.0329	0.132	0.538	2.200
`SimpleFloatArrayV`	0.00241	0.00948	0.0379	0.152	0.613	2.511
`ListFloatArray`	1.846	14.2928	—	—	—	—

Bemerkung:

- Die ersten fünf Zeilen zeigen deutlich den $O(n^2)$-Aufwand: Verdopplung von n bedeutet vierfache Laufzeit.

- Die Zeilen fünf und vier zeigen die Laufzeit für die Variante mit einem virtuellem `operator[]` bzw. eine Version der Klassenschablone, bei der das Schlüsselwort `inline` vor der Methodendefinition des `operator[]` weggelassen wurde. Diese beiden Varianten sind etwa viermal langsamer als die vorherigen.

- Eine Variante mit eingebautem Feld (nicht vorgestellt, ohne Klassen) ist am schnellsten, gefolgt von den zwei Varianten mit Klassenschablonen, die unwesentlich langsamer sind.

- Beachte auch den Einfluss des Optimizers (gcc-Option `-O3`).

- `ListFloatArray` ist die listenbasierte Darstellung des Feldes mit Index-Wert-Paaren. Diese hat Komplexität $O(n^3)$, da nun die Zugriffe auf die Feldelemente Komplexität $O(n)$ haben.

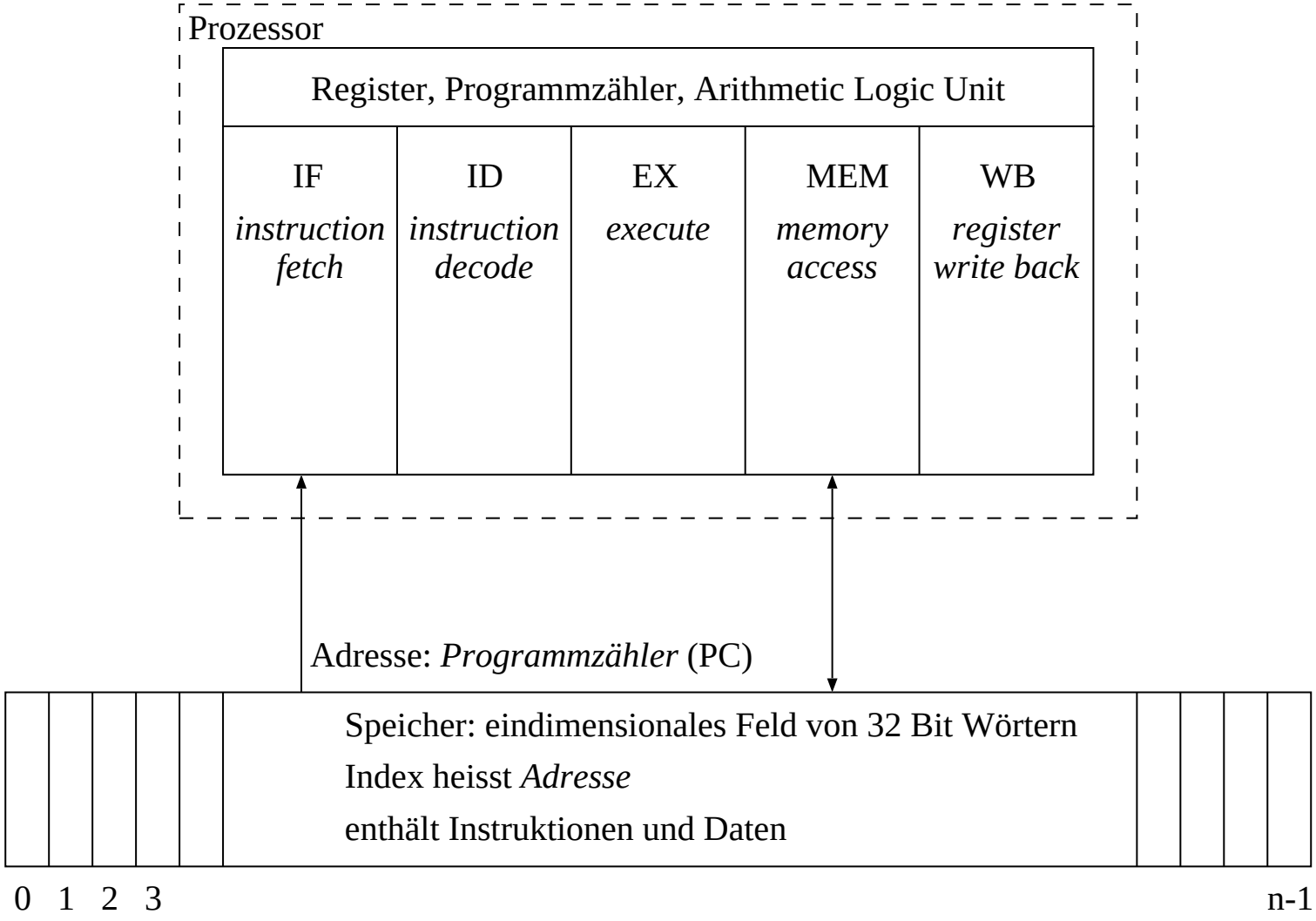
Frage: Warum sind die Varianten auf Schablonenbasis (mit inlining) schneller als die Variante mit virtueller Methode?

RISC

Bezeichnung: RISC steht für *Reduced Instruction Set Computer* und steht für eine Kategorie von Prozessorarchitekturen mit verhältnismäßig einfachem Befehlssatz. Gegenpol: CISC=*Complex Instruction Set Computer*.

Geschichte: RISC stellt heutzutage den Großteil aller Prozessoren dar (vor allem bei eingebetteten Systemen (Handy, PDA, Spielekonsole, etc.), wo das Verhältnis Leistung/Verbrauch wichtig ist). Für PCs ist allerdings noch mit den Intel-Chips die CISC-Technologie dominant (mittlerweile wurden aber auch dort viele RISC-Techniken übernommen).

Aufbau eines RISC-Chips


Prozessor
Register, Programmzähler, Arithmetic Logic Unit
IF
instruction fetch
ID
instruction decode
EX
execute
MEM
memory access
WB
register write back
Adresse: Programmzähler (PC)
Speicher: eindimensionales Feld von 32 Bit Wörtern
Index heisst Adresse
enthält Instruktionen und Daten
0 1 2 3
n-1

Befehlszyklus

Bezeichnung: Ein typischer RISC-Befehl lässt sich in Teilschritte unterteilen, die von verschiedener Hardware (in der CPU) ausgeführt werden:

1. IF: Holen des nächsten Befehls aus dem Speicher. Ort: Programmzähler.
2. ID: Dekodieren des Befehls, Auslesen der beteiligten Register.
3. EX: Eigentliche Berechnung (z. B. Addieren zweier Zahlen).
4. MEM: Speicherzugriff (entweder Lesen oder Schreiben).
5. WB: Rückschreiben der Ergebnisse in Register.

Dies nennt man Befehlszyklus (*instruction cycle*).

Pipelining

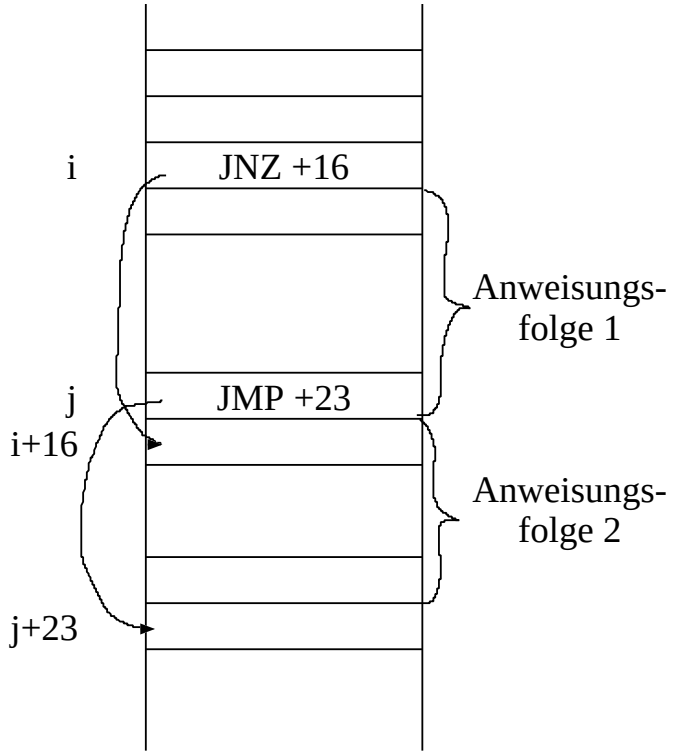
Diese Stadien werden nun für aufeinanderfolgende Befehle überlappend ausgeführt (Pipelining).

IF	Instr1	Instr2	Instr3	Instr4	Instr5	Instr6	Instr7
ID	—	Instr1	Instr2	Instr3	Instr4	Instr5	Instr6
EX	—	—	Instr1	Instr2	Instr3	Instr4	Instr5
MEM	—	—	—	Instr1	Instr2	Instr3	Instr4
WB	—	—	—	—	Instr1	Instr2	Instr3

Probleme mit Pipelining

Sehen wir uns an, wie eine `if`-Anweisung realisiert wird:

```
if ( a == 0 )
{
  <Anweisungsfolge 1>
}
else
{
  <Anweisungsfolge 2>
}
```



Problem: Das Sprungziel des Befehls `JNZ +16` steht erst am Ende der dritten Stufe der Pipeline (EX) zur Verfügung, da ein Register auf 0 getestet und 16 auf

den PC addiert werden muss.

Frage: Welche Befehle sollen bis zu diesem Punkt weiter angefangen werden?

Antwort:

- Gar keine, dann bleiben einfach drei Stufen der Pipeline leer (pipeline stall).

- Man $rät$ das Sprungziel (branch prediction unit) und führt die nachfolgenden Befehle spekulativ aus (ohne Auswirkung nach aussen). Notfalls muss man die Ergebnisse dieser Befehle wieder verwerfen.

Bemerkung: Selbst das Ziel eines unbedingten Sprungbefehls stünde wegen der Addition des Offset auf den PC erst nach der Stufe EX zur Verfügung (es sei denn, man hat extra Hardware dafür).

Funktionsaufrufe

Ein Funktionsaufruf (Methodenaufruf) besteht aus folgenden Operationen:

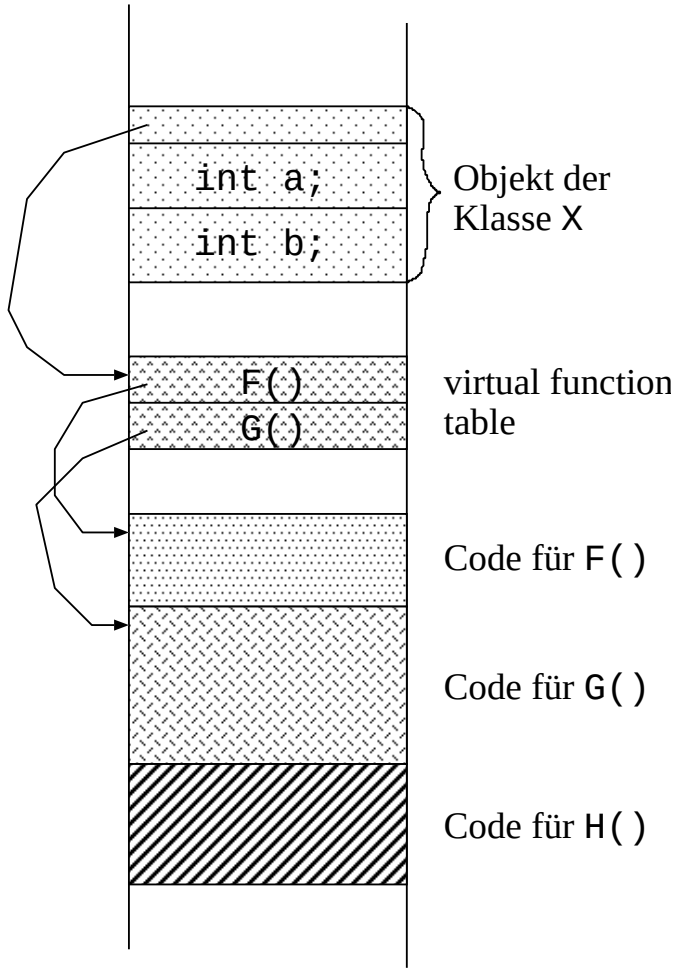
- Sicherung der Rücksprungadresse auf dem Stack
- ein unbedingter Sprungbefehl
- der Rücksprung an die gespeicherte Adresse
- + eventuelle Sicherung von Registern auf dem Stack

Diese Liste gilt genauso für CISC-Architekturen. Ein Funktionsaufruf ist also normalerweise mit erheblichem Aufwand verbunden. Darüberhinaus optimiert der Compiler nicht über Funktionsaufrufe hinweg, was zu weiteren Geschwindigkeitseinbussen führt.

Realisierung virtueller Funktionen

Betrachte folgende Klassendefinition und ein konkretes Objekt im Speicher:

```
class X
{
public:
   int a;
   int b;
   virtual void F();
   virtual void G();
   void H();
};

X x;
x.F();
x.H();
```



Bemerkung:

- Für jede Klasse gibt es eine Tabelle mit Zeigern auf den Programmcode für die virtuellen Funktionen dieser Klasse. Diese Tabelle heißt *virtual function table* (VFT).

- Jedes Objekt einer Klasse, die virtuelle Funktionen enthält, besitzt einen Zeiger auf die VFT der zugehörigen Klasse. Dies entspricht im wesentlichen der Typinformation, die bei Sprachen mit dynamischer Typbindung den Daten hinzugefügt ist.

- Beim Aufruf einer virtuellen Methode generiert der Übersetzer Code, welcher der VFT des Objekts die Adresse der aufzurufenden Methode entnimmt und dann den Funktionsaufruf durchführt. Welcher Eintrag der VFT zu entnehmen ist, ist zur Übersetzungszeit bekannt.

- Der Aufruf nichtvirtueller Funktionen geschieht ohne VFT. Klassen (und ihre

zugehörigen Objekte) ohne virtuelle Funktionen brauchen keinen Zeiger auf eine VFT.

- Für den Aufruf virtueller Funktionen ist immer ein Funktionsaufruf notwendig, da erst zur Laufzeit bekannt ist, welche Methode auszuführen ist.

Inlining

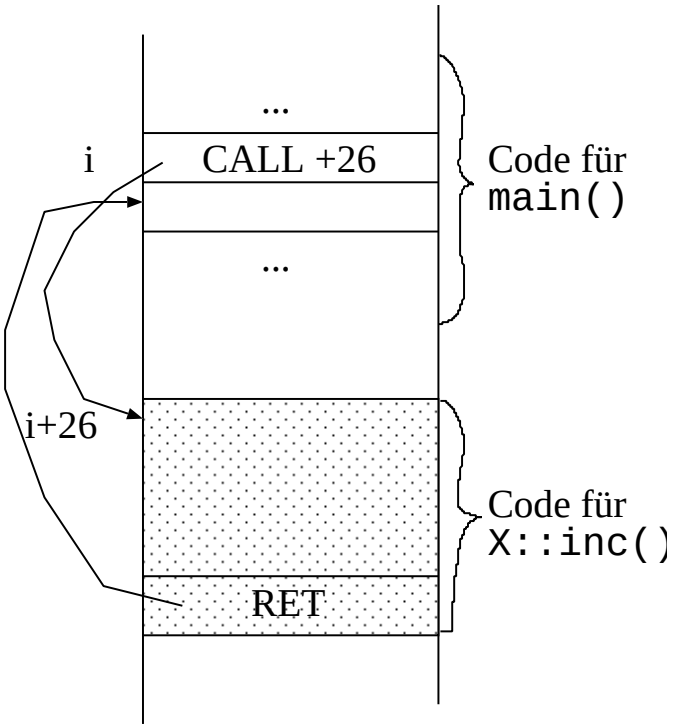
Problem: Der Funktionsaufruf sehr kurzer Funktionen ist relativ langsam.

Beispiel:

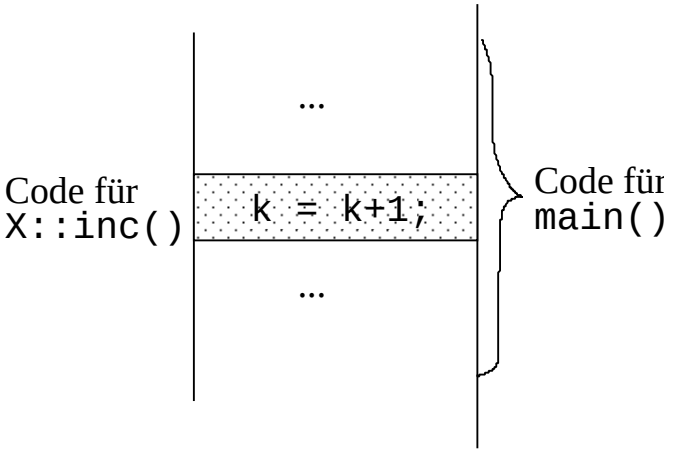
```
class X {
public:
  void inc();
private:
  int k;
};

inline void X::inc()
{
  k = k + 1;
}
```

```
void main()
{
  X x;
  x.inc();
}
```

Ohne das Schlüsselwort `inline` in der Methodendefinition generiert der C++-Übersetzer einen Funktionsaufruf für `inc()`:



Mit dem Schlüsselwort `inline` in der Methodendefinition setzt der Übersetzer den Code der Methode am Ort des Aufrufes direkt ein falls dies möglich ist:



Bemerkung:

- Inlining ändert nichts an der Semantik des Programmes.
- Das Schlüsselwort `inline` ist nur ein *Vorschlag* an den Compiler. Z. B. wird es für rekursive Funktionen ignoriert.
- Virtuelle Funktionen können nicht inline ausgeführt werden, da die auszuführende Methode zur Übersetzungszeit nicht bekannt ist.
- Aber: Änderungen der Implementation einer Inline-Funktion in einer Bibliothek machen normalerweise die erneute Übersetzung von anderen Programmteilen notwendig!

Bemerkung: Es sei auch nochmal eindringlich an Knuth's Wort *"Premature optimization is the root of all evil"* erinnert. Bevor Sie daran gehen, Ihr Programm durch Elimination virtueller Funktionen und Inlining unflexibler zu machen, sollten Sie folgendes tun:

1. Überdenken Sie den Algorithmus!

2. Messen Sie, wo der „Flaschenhals" wirklich liegt (Profiling notwendig).

3. Überlegen Sie, ob die erreichbare Effizienzsteigerung den Aufwand wert ist.

Beispielsweise ist die einzig sinnvolle Verbesserung für das Sortierbeispiel am Anfang dieses Abschnitts das Verwenden eines besseren Algorithmus!

Zusammenfassung

- Klassenschablonen definieren parametrisierte Datentypen und sind daher besonders geeignet, um allgemein verwendbare Konzepte (ADT) zu implementieren.
- Funktionsschablonen definieren parametrisierte Funktionen, die auf verschiedenen Datentypen (mit gleicher Schnittstelle) operieren.
- In beiden Fällen werden konkrete Varianten der Klassen/Funktionen zur Übersetzungszeit erzeugt und übersetzt (generische Programmierung).
- Diese Techniken sind für Sprachen mit dynamischer Typbindung meist unnötig. Solche Sprachen brauchen aber in vielen Fällen Typabfragen zur Laufzeit, was dazu führt, dass der erzeugte Code nicht mehr hocheffizient ist.

Nachteile der generischen Programmierung

- Es wird viel Code erzeugt. Die Übersetzungszeiten template-intensiver Programme können unerträglich lang sein.

- Es ist keine getrennte Übersetzung möglich. Der Übersetzer muss die Definition aller vorkommenden Schablonen kennen. Dasselbe gilt für Inline-Funktionen. Dies erfordert dann z. B. auch spezielle Softwarelizenzen.

- Das Finden von Fehlern in Klassen/Funktionenschablonen ist erschwert, da der Code für eine konkrete Variante nirgends existiert. Empfehlung: testen Sie zuerst mit einem konkreten Datentyp und machen Sie dann eine Schablone daraus.

Containerklassen

Motivation

Bezeichnung: Klassen, die eine Menge anderer Objekte verwalten (man sagt aggregieren) nennt man Containerklassen.

Beispiele: Wir hatten: Liste, Stack, Feld. Weitere sind: binärer Baum (*binary tree*), Warteschlange (*queue*), Abbildung (*map*), . . .

Bemerkung: Diese Strukturen treten sehr häufig als Komponenten in größeren Programmen auf. Ziel von Containerklassen ist es, diese Bausteine in wiederverwendbarer Form zur Verfügung zu stellen (*code reuse*).

Vorteile:

- Weniger Zeitaufwand in Entwicklung und Fehlersuche.
- Klarere Programmstruktur, da man auf einer höheren Abstraktionsebene arbeitet.

Werkzeug: Das Werkzeug zur Realisierung effizienter und flexibler Container in C++ sind Klassenschablonen.

Bemerkung: In diesem Abschnitt sehen wir uns eine Reihe von Containern an. Die Klassen sind vollständig ausprogrammiert und zeigen, wie man Container implementieren *könnte*. In der Praxis verwendet man allerdings die *Standard Template Library* (STL), welche Container in professioneller Qualität bereitstellt.

Ziel: Sie sind am Ende dieses Kapitels motiviert, die STL zu verwenden und können die Konzepte verstehen.

Listenschablone

Bei diesem Entwurf ist die Idee, das Listenelement und damit auch die Liste als Klassenschablone zu realisieren. In jedem Listenelement wird ein Objekt der Klasse `T`, dem Schablonenparameter, gespeichert.

Programm: Definition und Implementation (Liste.hh)

```
template <class T>
class List
{
public:
  // Infrastruktur
  List() { _first = 0; }
  ~List();

  // Listenelement als nested class
  class Link
  {
```

```
    Link* _next;
  public:
    T item;
    Link( T& t ) { item = t; }
    Link* next() { return _next; }
    friend class List<T>;
  };

  Link* first() { return _first; }
  void insert( Link* where, T t );
  void remove( Link* where );

private:
  Link* _first;
  // privater Copy-Konstruktor und Zuweisungs-
  // operator da Defaultvarianten zu fehlerhaftem
  // Verhalten fuehren
  List( const List<T>& l ) {};
  List<T>& operator=( const List<T>& l ) {};
```

```
};

template <class T> List<T>::~List()
{
  Link* p = _first;
  while ( p != 0 )
  {
    Link* q = p;
    p = p->next();
    delete q;
  }
}

template <class T>
void List<T>::insert( List<T>::Link* where, T t )
{
  Link* ins = new Link(t);
  if ( where == 0 )
  {
```

```
    ins->_next = _first;
    _first = ins;
  }
  else
  {
    ins->_next = where->_next;
    where->_next = ins;
  }
}

template <class T>
void List<T>::remove( List<T>::Link* where )
{
  Link* p;
  if ( where == 0 )
  {
    p = _first;
    if ( p != 0 ) _first = p->_next;
  }
```

```
    else
    {
      p = where->_next;
      if ( p != 0 ) where->_next = p->_next;
    }
    delete p;
}
```

Programm: Verwendung (UseListe.cc)

```
#include <iostream>
#include "Liste.hh"

int main()
{
  List<int> list;

  list.insert( 0, 17 );
  list.insert( 0, 34 );
  list.insert( 0, 26 );

  for ( List<int>::Link* l=list.first(); l!=0; l=l->next() )
    std::cout << l->item << std::endl;
  for ( List<int>::Link* l=list.first(); l!=0; l=l->next() )
    l->item = 23;
}
```

Bemerkung:

- Diese Liste ist homogen, d. h. alle Objekte im Container haben den gleichen Typ. Eine heterogene Liste könnte man als Liste von Zeigern auf eine gemeinsame Basisklasse realisieren.

- Speicherverwaltung wird von der Liste gemacht. Listen können kopiert und als Parameter übergeben werden (sofern Copy-Konstruktor und Zuweisungsoperator noch mittels deep copy implementiert werden).

- Zugriff auf die Listenelemente erfolgt über eine offengelegte *nested class*. Die Liste wird als friend deklariert, damit die Liste den `next`-Zeiger manipulieren kann, nicht jedoch der Benutzer der Liste.

Iteratoren

Problem: Grundoperation aller Container sind

- Durchlaufen aller Elemente,
- Zugriff auf Elemente.

Um Container austauschbar verwenden zu können, sollten diese Operationen mit der gleichen Schnittstelle möglich sein. Die Schleife für eine Liste sah aber ganz anders aus als bei einem Feld.

Abhilfe: Diese Abstraktion realisiert man mit Iteratoren. Iteratoren sind zeigerähnliche Objekte, die auf ein Objekt im Container zeigen (obwohl der Iterator nicht als Zeiger realisiert sein muss).

Prinzip:

```
template <class T> class Container
{
public:
  class Iterator
  {  // nested class definition
    ...
  public:
    Iterator();
    bool operator!=( Iterator x );
    bool operator==( Iterator x );
    Iterator operator++();        // prefix
    Iterator operator++( int );   // postfix
    T& operator*() const;
    T* operator->() const;
    friend class Container<T>;
  };
  Iterator begin() const;
  Iterator end() const;
```

```
  ... // Spezialitäten des Containers
};

// Verwendung
Container<int> c;
for ( Container<int>::Iterator i=c.begin(); i!=c.end(); ++i )
  std::cout << *i << std::endl;
```

Bemerkung:

- Der `Iterator` ist als Klasse innerhalb der Containerklasse definiert. Dies nennt man eine geschachtelte Klasse (*nested class*).
- Damit drückt man aus, dass `Container` und `Iterator` zusammengehören. Jeder Container wird seine eigene Iteratorklasse haben.
- Innerhalb von `Container` kann man `Iterator` wie jede andere Klasse verwenden.

- `friend class Container<T>` bedeutet, dass die Klasse `Container<T>` auch Zugriff auf die *privaten* Datenmitglieder der Iteratorklasse hat.

- Die Methode `begin()` des Containers liefert einen Iterator, der auf das erste Element des Containers zeigt.

- `++i` bzw. `i++` stellt den Iterator auf das $n\ddot{a}chste$ Element im Container. Zeigte der Iterator auf das letzte Element, dann ist der Iterator gleich dem von `end()` gelieferten Iterator.

- `++i` bzw. `i++` manipulieren den Iterator für den sie aufgerufen werden. Als Rückgabewert liefert `++i` den neuen Wert des Iterators, `i++` jedoch den alten.

- Bei der Definition unterscheiden sie sich dadurch, dass der Postfix-Operator noch ein **int**-Argument erhält, das aber keine Bedeutung hat.

- `end()` liefert einen Iterator, der auf „das Element nach dem letzten Element“ des Containers zeigt (siehe oben).

- `*i` liefert eine Referenz auf das Objekt im Container, auf das der Iterator `i` zeigt. Damit kann man sowohl `x = *i` als auch `*i = x` schreiben.

- Ist das Objekt im Container von einem zusammengesetzten Datentyp (also `struct` oder `class`), so kann mittels `i-><Komponente>` eine Komponente selektiert werden. Der Iterator verhält sich also wie ein Zeiger.

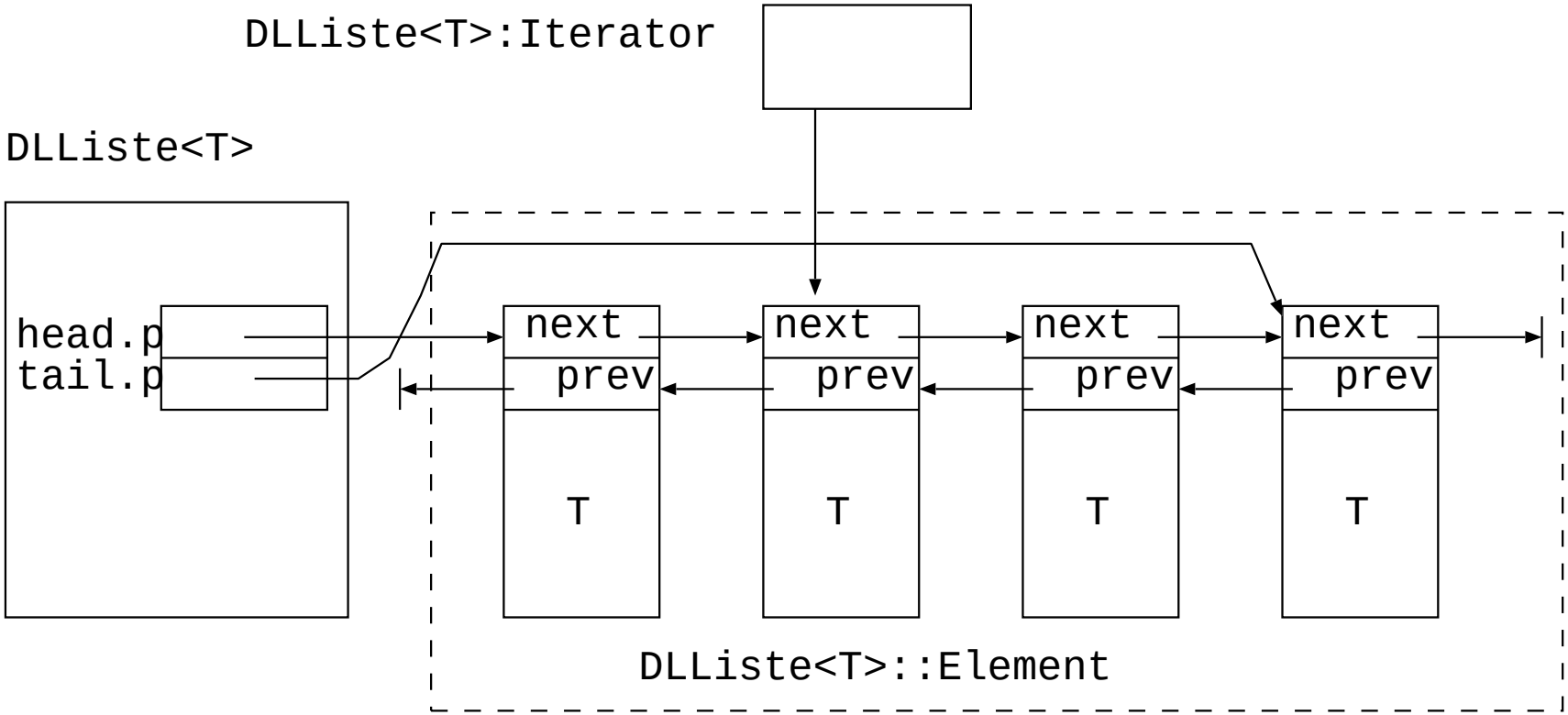
Machen wir ein Beispiel . . .

Doppelt verkettete Liste

Anforderungen:

- Vorwärts- und Rückwärtsdurchlauf
- Das Einfügen vor oder nach einem Element soll eine $O(1)$-Operation sein. Die Position wird durch einen Iterator angegeben.
- Das Entfernen eines Elementes soll eine $O(1)$-Operation sein. Das zu entfernende Element wird wieder durch einen Iterator angegeben
- Für die Berechnung der Größe der Liste akzeptieren wir einen $O(N)$-Aufwand.

Struktur


DLListe<T>:Iterator
DLListe<T>
head.p
tail.p
next
prev
T
next
prev
T
next
prev
T
next
prev
T
DLListe<T>::Element

Bemerkung:

- Intern werden die Listenelemente durch den Datentyp `Element` repräsentiert. Dieser private, geschachtelte, zusammengesetzte Datentyp ist außerhalb der Klasse nicht sichtbar.

- Die Einfügeoperationen erhalten Objekte vom Typ `T`, erzeugen dynamisch ein Listenelement und $kopieren$ das Objekt in das Listenelement.

- Damit kann man Listen für beliebige Datentypen erzeugen. Die Manipulation gelingt mit Hilfe der Iteratorschnittstelle. Der Iterator kapselt insbesondere den Zugriff auf die außerhalb der Liste nicht bekannten `Element`-Objekte.

Implementation

Programm: DLL.hh

```
template <class T>
class DLList
{
  // das interne Listenelement
  struct Element
  {
    Element* next;
    Element* prev;
    T item;
    Element( T &t )
    {
      item = t;
      next = prev = 0;
    }
  };
```

```
public:
  typedef T MemberType;   // Merke Grundtyp

  // der iterator kapselt Zeiger auf Listenelement
  class Iterator
  {
  private:
    Element* p;
  public:
    Iterator() { p = 0; }
    Iterator( Element* q ) { p = q; }
    bool operator!=( Iterator x ) {
      return p != x.p;
    }
    bool operator==( Iterator x ) {
      return p == x.p;
    }
    Iterator operator++() { // prefix
```

```
    p = p->next;
    return *this;
  }
  Iterator operator++( int ) { // postfix
    Iterator tmp = *this;
    p = p->next;
    return tmp;
  }
  Iterator operator--() { // prefix
    p = p->prev;
    return *this;
  }
  Iterator operator--( int ) { // postfix
    Iterator tmp = *this;
    p = p->prev;
    return tmp;
  }
  T& operator*() { return p->item; }
  T* operator->() { return &(p->item); }
```

```
  friend class DLList<T>; // Liste manip. p
};

// Iteratoren
Iterator begin() const { return head; }
Iterator end() const { return Iterator(); }
Iterator rbegin() const { return tail; }
Iterator rend() const { return Iterator(); }

// Konstruktion, Destruktion, Zuweisung
DLList();
DLList( const DLList<T>& list );
DLList<T>& operator=( const DLList<T>& );
~DLList();

// Listenmanipulation
Iterator insert( Iterator i, T t ); // einf. vor i
void erase( Iterator i );
void append( const DLList<T>& l );
```

```
  void clear();
  bool empty() const;
  int size() const;
  Iterator find( T t ) const;

private:
  Iterator head;     // erstes Element der Liste
  Iterator tail;     // letztes Element der Liste
};


// Insertion
template <class T>
typename DLList<T>::Iterator
DLList<T>::insert( Iterator i, T t )
{
  Element* e = new Element( t );
  if ( empty() )
  {
```

```
    assert( i.p == 0 );
    head.p = tail.p = e;
  }
  else
  {
    e->next = i.p;
    if ( i.p != 0 )
    {    // insert before i
      e->prev = i.p->prev;
      i.p->prev = e;
      if ( head == i )
        head.p = e;
    }
    else
    {    // insert at end
      e->prev = tail.p;
      tail.p->next = e;
      tail.p = e;
    }
```

```
  }
  return Iterator( e );
}

template <class T>
void DLList<T>::erase( Iterator i )
{
  if ( i.p == 0 ) return;

  if ( i.p->next != 0 )
    i.p->next->prev = i.p->prev;
  if ( i.p->prev != 0 )
    i.p->prev->next = i.p->next;

  if ( head == i ) head.p = i.p->next;
  if ( tail == i ) tail.p = i.p->prev;

  delete i.p;
}
```

```
template <class T>
void DLList<T>::append( const DLList<T>& l ) {
  for ( Iterator i=l.begin(); i!=l.end(); i++ )
    insert( end(), *i );
}

template <class T>
bool DLList<T>::empty() const {
  return begin() == end();
}

template <class T>
void DLList<T>::clear() {
  while ( !empty() )
    erase( begin() );
}

// Constructors
```

```
template <class T> DLList<T>::DLList() {}

template <class T>
DLList<T>::DLList( const DLList<T>& list ) {
  append( list );
}

// Assignment
template <class T>
DLList<T>&
DLList<T>::operator=( const DLList<T>& l )
{
  if ( this != &l )
  {
    clear();
    append(l);
  }
  return *this;
}
```

```
// Destructor
template <class T> DLList<T>::~DLList() { clear(); }

// Size method
template <class T> int DLList<T>::size() const
{
  int count = 0;
  for ( Iterator i=begin(); i!=end(); i++ )
    count++;
  return count;
}

template <class T>
typename DLList<T>::Iterator DLList<T>::find( T t ) const
{
  DLList<T>::Iterator i = begin();
  while ( i != end() )
  {
```

```
    if ( *i == t ) break;
    i++;
  }
  return i;
}

template <class T>
std::ostream& operator<<( std::ostream& s, DLList<T>& a )
{
  s << "(";
  for ( typename DLList<T>::Iterator i=a.begin();
        i!=a.end(); i++ )
  {
    if ( i != a.begin() ) s << " ";
    s << *i;
  }
  s << ")" << std::endl;
  return s;
}
```

Verwendung

Programm: UseDLL.cc

```
#include <cassert>
#include <iostream>
#include "DLL.hh"
#include "Zufall.cc"

int main()
{
  Zufall z( 87124 );
  DLList<int> l1, l2, l3;

  // Erzeuge 3 Listen mit je 5 Zufallszahlen
  for ( int i=0; i<5; i=i+1 )
    l1.insert( l1.end(), i );
  for ( int i=0; i<5; i=i+1 )
    l2.insert( l2.end(), z.ziehe_zahl() );
```

```
for ( int i=0; i<5; i=i+1 )
  l3.insert( l3.end(), z.ziehe_zahl() );

// Loesche alle geraden in der ersten Liste
DLList<int>::Iterator i, j;
i = l1.begin();
while ( i != l1.end() )
{
  j = i; // merke aktuelles Element
  ++i;   // gehe zum naechsten
  if ( *j % 2 == 0 ) l1.erase( j );
}

// Liste von Listen ...
DLList<DLList<int> > ll;
ll.insert( ll.end(), l1 );
ll.insert( ll.end(), l2 );
ll.insert( ll.end(), l3 );
std::cout << ll << std::endl;
```

```
  std::cout << "Laenge: " << ll.size() << std::endl;
}
```

Diskussion

- Den Rückwärtsdurchlauf durch eine Liste c erreicht man durch:

```
for ( DLList<int>::Iterator i=c.rbegin();
      i!=c.rend(); --i )
  std::cout << *i << endl;
```

- Die Objekte (vom Typ `T`) werden beim Einfügen in die Liste kopiert. Abhilfe: Liste von Zeigern auf die Objekte, z. B. `DLList<int*>`.

- Die Schlüsselworte **const** in der Definition von `begin`, `end`, ... bedeuten, dass diese Methoden ihr Objekt nicht ändern.

- Innerhalb einer Template-Definition werden geschachtelte Klassen nicht als Typ erkannt. Daher muss man den Namen explizit mittels **typename** als Typ kennzeichnen.

Beziehung zur STL-Liste

Die entsprechende STL-Schablone heißt `list` und unterscheidet sich von unserer Liste unter anderem in folgenden Punkten:

- Man erhält die Funktionalität durch **#include** <list>.
- Die Iterator-Klasse heißt `iterator` statt `Iterator`.
- Es gibt zusätzlich einen `const_iterator`. Auch unterscheiden sich Vorwärts- und Rückwärtsiteratoren (`reverse_iterator`).
- Sie hat einige Methoden mehr, z. B. `push_front`, `push_back`, `front`, `back`, `pop_front`, `pop_back`, `sort`, `reverse`, . . .
- Die Ausgabe über „`std::cout` <<“ ist nicht definiert.

Feld

Wir fügen nun die Iterator-Schnittstelle unserer `SimpleArray<T>`-Schablone hinzu.

Programm: (Array.hh)

```
template <class T> class Array
{
public:
  typedef T MemberType;   // Merke Grundtyp

  // Iterator fuer die Feld-Klasse
  class Iterator
  {
  private:
    T* p; // Iterator ist ein Zeiger ...
    Iterator( T* q ) { p = q; }
  public:
    Iterator() { p = 0; }
    bool operator!=( Iterator x ) {
```

```
    return ( p != x.p );
  }
  bool operator==( Iterator x ) {
    return ( p == x.p );
  }
  Iterator operator++() {
    p++;
    return *this;
  }
  Iterator operator++( int ) {
    Iterator tmp = *this;
    ++*this;
    return tmp;
  }
  T& operator*() const { return *p; }
  T* operator->() const { return p; }
  friend class Array<T>;
};
```

```
// Iterator Methoden
Iterator begin() const {
  return Iterator( p );
}
Iterator end() const {
  return Iterator( &(p[n]) ); // ja, das ist ok!
}

// Konstruktion; Destruktion und Zuweisung
Array( int m ) {
  n = m;
  p = new T[n];
}
Array( const Array<T>& );
Array<T>& operator=( const Array<T>& );
~Array() {
  delete[] p;
}
```

```
  // Array manipulation
  int size() const {
    return n;
  }
  T& operator[]( int i ) {
    return p[i];
  }

private:
  int n;   // Anzahl Elemente
  T* p;    // Zeiger auf built-in array
};

// Copy-Konstruktor
template <class T>
Array<T>::Array( const Array<T>& a ) {
  n = a.n;
  p = new T[n];
  for ( int i=0; i<n; i=i+1 )
```

```
    p[i] = a.p[i];
}

// Zuweisung
template <class T>
Array<T>& Array<T>::operator=( const Array<T>& a ) {
  if ( &a != this )
  {
    if ( n != a.n )
    {
      delete[] p;
      n = a.n;
      p = new T[n];
    }
    for ( int i=0; i<n; i=i+1 ) p[i] = a.p[i];
  }
  return *this;
}
```

```
// Ausgabe
template <class T>
std::ostream& operator<<( std::ostream& s, Array<T>& a ) {
  s << "array " << a.size() << " elements = [" << std::endl;
  for ( int i=0; i<a.size(); i++ )
    s << "    " << i << "  " << a[i] << std::endl;
  s << "]" << std::endl;
  return s;
}
```

Bemerkung:

- Der Iterator ist als Zeiger auf ein Feldelement realisiert.
- Die Schleife

  ```
  for (Array<int>::Iterator i=a.begin(); i!=a.end(); ++i) ...
  ```

  entspricht nach Inlining der Methoden einfach

  ```
  for (int* p=a.p; p!=&a[100]; p=p+1) ...
  ```

  und ist somit nicht langsamer als handprogrammiert!
- Man beachte auch die Definition von `MemberType`. Dies ist praktisch innerhalb eines Template **template** <**class** C>, wo der Datentyp eines Containers `C` dann als `C::MemberType` erhalten werden kann.

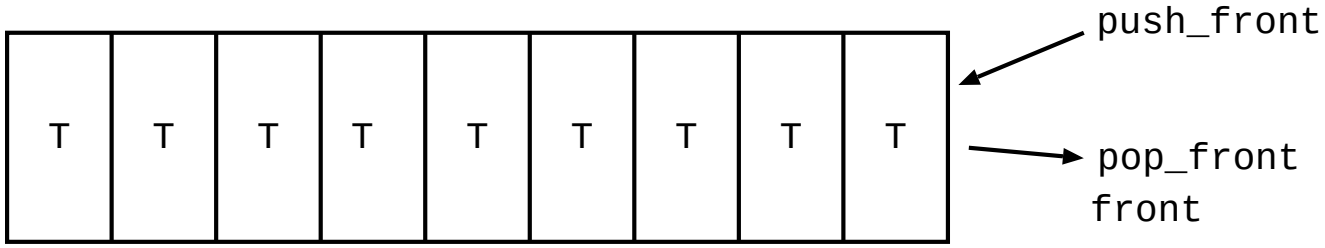
Programm: **Gleichzeitige Verwendung** `DLList/Array` **(UseBoth.cc)**:

```
#include <cassert>
#include <iostream>

#include "Array.hh"
#include "DLL.hh"
#include "Zufall.cc"

int main()
{
  Zufall z( 87124 );
  Array<int> a( 5 );
  DLList<int> l;

  // Erzeuge Array und Liste mit 5 Zufallszahlen
  for ( int i=0; i<5; i=i+1 ) a[i] = z.ziehe_zahl();
  for ( int i=0; i<5; i=i+1 )
    l.insert( l.end(), z.ziehe_zahl() );
```

```
  // Benutzung
  for ( Array<int>::Iterator i=a.begin(); i!=a.end(); i++ )
    std::cout << *i << std::endl;
  std::cout << std::endl;
  for ( DLList<int>::Iterator i=l.begin(); i!=l.end(); i++ )
    std::cout << *i << std::endl;
}
```

Bemerkung: Die STL-Version von `Array` erhält man mit **#include** <vector>. Die Klassenschablone heißt `vector` anstatt `Array`.

Stack



Schnittstelle:

- Konstruktion eines Stack.
- Einfügen eines Elementes vom Typ `T` oben (`push`).
- Entfernen des obersten Elementes (`pop`).
- Inspektion des obersten Elementes (`top`).
- Test ob Stack voll oder leer (`empty`).

Programm: Implementation über DLList (Stack.hh)

```
template <class T>
class Stack : private DLList<T>
{
public:
  // Default-Konstruktoren + Zuweisung OK

  bool empty() { return DLList<T>::empty(); }
  void push( T t ) {
    insert( begin(), t );
  }
  T top() { return *begin(); }
  void pop() { erase( begin() ); }
};
```

Bemerkung:

- Wir haben den Stack als Spezialisierung der doppelt verketteten Liste realisiert. Etwas effizienter wäre die Verwendung einer einfach verketteten Liste gewesen.

- Auffallend ist, dass die Befehle `top`/`pop` getrennt existieren (und `pop` keinen Wert zurückliefert). Verwendet werden diese Befehle nämlich meist gekoppelt, so dass auch eine Kombination `pop` $\leftarrow$ `top`+`pop` nicht schlecht wäre.

Programm: **Anwendung: (UseStack.cc)**

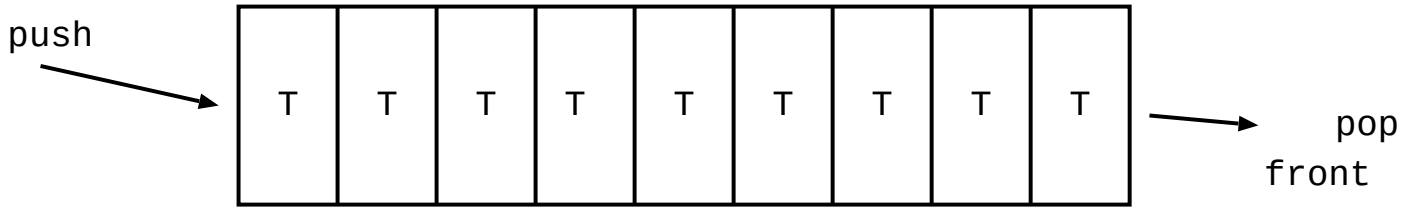
```
#include <cassert>
#include <iostream>

#include "DLL.hh"
#include "Stack.hh"

int main()
{
  Stack<int> s1;
  for ( int i=1; i<=5; i++ )
    s1.push( i );

  Stack<int> s2( s1 );
  s2 = s1;
  while ( !s2.empty() )
```

```
    {
        std::cout << s2.top() << std::endl;
        s2.pop();
    }
}
```

Bemerkung: Die STL-Version erhält man durch **#include** <stack>. Die Klassenschablone heißt dort `stack` und hat im wesentlichen dieselbe Schnittstelle.

Queue

Eine Queue ist eine Struktur, die Einfügen an einem Ende und Entfernen nur am anderen Ende erlaubt:



Anwendung: Warteschlangen.

Schnittstelle:

- Konstruktion einer leeren Queue
- Einfügen eines Elementes vom Typ `T` am Ende
- Entfernen des Elementes am Anfang
- Inspektion des Elementes am Anfang
- Test ob Queue leer

Programm: **(Queue.hh)**

```
template <class T>
class Queue : private DLList<T>
{
public:
  // Default-Konstruktoren + Zuweisung OK
  bool empty() {
    return DLList<T>::empty();
  }
  T front() {
    return *DLList<T>::begin();
  }
  T back() {
    return *DLList<T>::rbegin();
  }
  void push( T t ) {
```

```
    insert( DLList<T>::end(), t );
  }
  void pop() {
    erase( DLList<T>::begin() );
  }
};
```

Bemerkung: Die STL-Version erhält man durch **#include** <queue>. Die Klassenschablone heißt dort queue und hat im wesentlichen dieselbe Schnittstelle wie Queue.

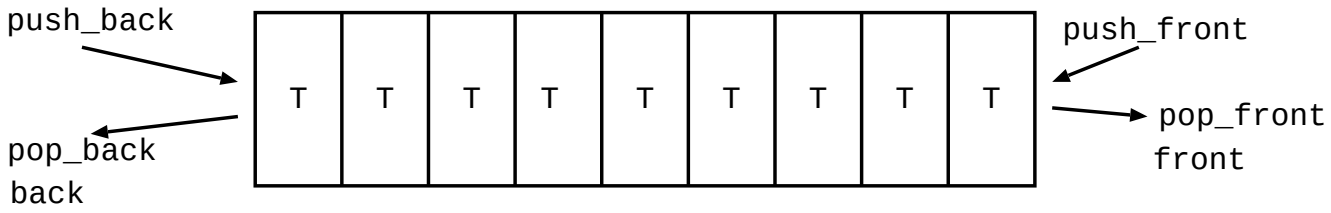
Programm: Zur Abwechslung verwenden wir mal die STL-Version: **(UseQueueSTL.cc)**

```
#include <queue>
#include <iostream>

int main()
{
  std::queue<int> q;
  for ( int i=1; i<=5; i++ )
    q.push( i );

  while ( !q.empty() )
  {
    std::cout << q.front() << std::endl;
    q.pop();
  }
```

```
}
```

DeQueue

Eine DeQueue (*double-ended queue*) ist eine Struktur, die Einfügen und Entfernen an beiden Enden erlaubt:



Schnittstelle:

- Konstruktion einer leeren `DeQueue`
- Einfügen eines Elementes vom Typ `T` am Anfang oder Ende
- Entfernen des Elementes am Anfang oder am Ende
- Inspektion des Elementes am Anfang oder Ende
- Test ob `DeQueue` leer

Programm: **(DeQueue.hh)**

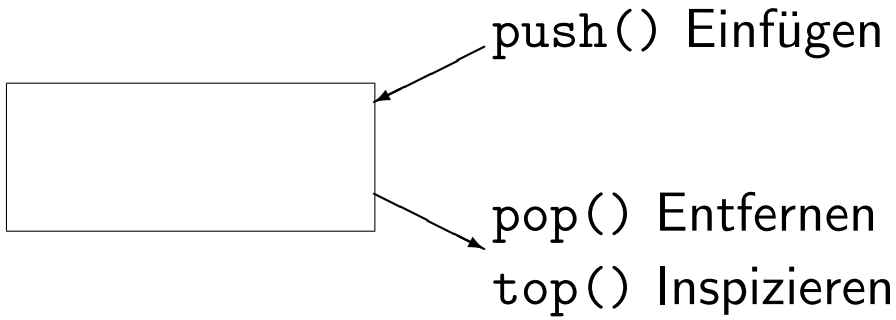
```
template <class T>
class DeQueue : private DLList<T>
{
public:
  // Default-Konstruktoren + Zuweisung ok
  bool empty();
  void push_front( T t );
  void push_back( T t );
  T pop_front();
  T pop_back();
  T front();
  T back();
};
```

Bemerkung: Die STL-Version erhält man auch hier mit **#include** <queue>. Die Klassenschablone heißt deque.

Prioritätswarteschlangen

Bezeichnung: Eine Prioritätswarteschlange ist eine Struktur, in die man Objekte des Grundtyps `T` einfüllen kann und von der jeweils das *kleinste* (MinPriorityQueue) bzw. das *größte* (MaxPriorityQueue) der eingegebenen Elemente als nächstes entfernt werden kann. Bei gleich großen Elementen verhält sie sich wie eine Queue.

Bemerkung: Auf dem Grundtyp `T` muß dazu die Relation $<$ mittels dem **operator<** zur Verfügung stehen.



Schnittstelle:

- Konstruktion einer leeren MinPriorityQueue.
- Einfügen eines Elementes vom Typ `T` (`push`).
- Entfernen des kleinsten Elementes im Container (`pop`).
- Inspektion des kleinsten Elementes im Container (`top`).
- Test ob MinPriorityQueue leer (`empty`).

Programm: Hier die Klassendefinition **(MinPriorityQueue.hh)**:

```
template <class T>
class MinPriorityQueue : private DLList<T>
{
private:
  typename DLList<T>::Iterator find_minimum();
public:
  // Default-Konstruktoren + Zuweisung OK
  bool empty();
  void push( T t );   // Einfuegen
  void pop();         // Entferne kleinstes
  T top();            // Inspiziere kleinstes
};
```

Und die Implementation **(MinPriorityQueueImp.cc)**:

```
template <class T>
bool MinPriorityQueue<T>::empty() {
  return DLList<T>::empty();
```

```
}

template <class T>
void MinPriorityQueue<T>::push( T t ) {
  insert( DLList<T>::begin(), t );
}

template <class T>
typename DLList<T>::Iterator
MinPriorityQueue<T>::find_minimum()
{
  typename DLList<T>::Iterator min = DLList<T>::begin();
  for ( typename DLList<T>::Iterator i=DLList<T>::begin();
        i!=DLList<T>::end(); i++ )
    if ( *i <= *min ) min = i;
  return min;
}

template <class T>
```

```
inline void MinPriorityQueue<T>::pop() {
  erase( find_minimum() );
}

template <class T>
inline T MinPriorityQueue<T>::top() {
  return *find_minimum();
}
```

Bemerkung:

- Unsere Implementierung arbeitet mit einer einfach verketteten Liste. Das Einfügen hat Komplexität $O(1)$, das Entfernen/Inspizieren jedoch $O(n)$.

- Bessere Implementationen verwenden einen Heap, was zu einem Aufwand der Ordnung $O(\log n)$ führt.

- Analog ist die Implementation der `MaxPriorityQueue`.

Bemerkung: Die STL-Version erhält man auch durch **#include** <queue>. Die Klassenschablone heißt `priority_queue` und implementiert eine `MaxPriorityQueue`. Man kann allerdings den Vergleichsoperator auch als Template-Parameter übergeben (etwas lästig).

Set

Ein Set (Menge) ist ein Container mit folgenden Operationen:

- Konstruktion einer leeren Menge.
- Einfügen eines Elementes vom Typ `T`.
- Entfernen eines Elementes.
- Test auf Enthaltensein.
- Test ob Menge leer.

Programm: Klassendefinition **(Set.hh)**:

```
template <class T>
class Set : private DLList<T>
{
public:
  // Default-Konstruktoren + Zuweisung OK

  typedef typename DLList<T>::Iterator Iterator;
  Iterator begin();
  Iterator end();

  bool empty();
  bool member( T t );
  void insert( T t );
  void remove( T t );
  // union, intersection, ... ?
};
```

Implementation **(SetImp.cc)**:

```
template <class T>
typename Set<T>::Iterator Set<T>::begin() {
  return DLList<T>::begin();
}

template <class T>
typename Set<T>::Iterator Set<T>::end() {
  return DLList<T>::end();
}

template <class T>
bool Set<T>::empty() {
  return DLList<T>::empty();
}

template <class T>
inline bool Set<T>::member( T t ) {
```

```
  return find( t ) != DLList<T>::end();
}

template <class T>
inline void Set<T>::insert( T t )
{
  if ( !member( t ) )
    DLList<T>::insert( DLList<T>::begin(), t );
}

template <class T>
inline void Set<T>::remove( T t )
{
  typename DLList<T>::Iterator i = find( t );
  if ( i != DLList<T>::end() )
    erase( i );
}
```

Bemerkung:

- Die Implementierung hier basiert auf der doppelt verketteten Liste von oben (private Ableitung!).
- Einfügen, Suchen und Entfernen hat die Komplexität $O(n)$.
- Wir lernen später Implementierungen kennen, die den Aufwand $O(\log n)$ für alle Operationen haben.
- Auf dem Typ `T` muss der Vergleichsoperator **operator**$<$ definiert sein. (Set gehört zu den sog. sortierten, assoziativen Containern).

Map

Bezeichnung: Eine Map ist ein assoziatives Feld, das Objekten eines Typs `Key` Objekte eines Typs `T` zuordnet.

Beispiel: Telefonbuch:

Meier	$\rightarrow$	504423
Schmidt	$\rightarrow$	162300
Müller	$\rightarrow$	712364
Huber	$\rightarrow$	8265498

Diese Zuordnung könnte man realisieren mittels:

```
Map<string, int> telefonbuch;
telefonbuch["Meier"] = 504423;
...
```

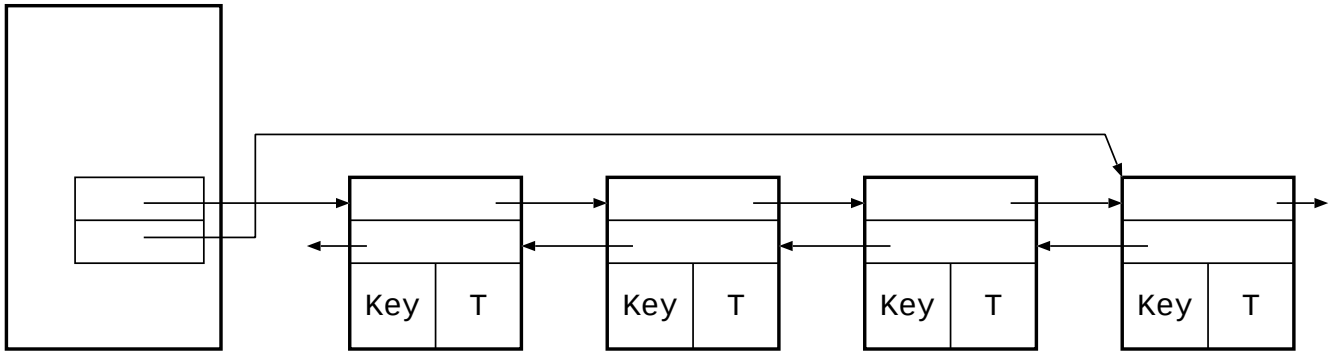
Programm: Definition der Klassenschablone (Map.hh)

```
// Existiert schon als std::pair
// template <class Key, class T>
// struct pair {
//     Key first;
//     T second;
// };

template <class Key, class T>
class Map : private DLList<pair<Key, T> >
{
public:

  T& operator[]( const Key& k );

  typedef typename DLList<pair<Key, T> >::Iterator Iterator;
  Iterator begin() const;
  Iterator end() const;
```

```
  Iterator find( const Key& k );
};
```

Bemerkung:

- In dieser Implementation von Map werden je zwei Objekte der Typen `Key` (*Schlüssel*) und `T` (*Wert*) zu einem Paar vom Typ `pair<Key,T>` kombiniert und in eine doppelt verkettete Liste eingefügt:



- Ein Objekt vom Typ `Key` kann nur einmal in einer Map vorkommen. (Daher ist das Telefonbuch kein optimales Beispiel.)

- Wir haben einen Iterator zum Durchlaufen des Containers.
- Auf dem Schlüssel muss der Vergleichsoperator **operator**$<$ definiert sein.
- `find(Key k)` liefert den Iterator für den Wert, ansonsten `end()`.
- Der Aufwand einer Suche ist wieder $O(n)$. Bald werden wir aber eine Realisierung von `Map` kennenlernen, die Einfügen und Suchen in $O(\log n)$ Schritten ermöglicht.

Wiederholung: Containerklassen

- Container speichern Objekte eines Grundtyps `T`.
- Container erlauben Inspektion der Elemente und Modifikation des Inhalts.
- Diese Operationen werden mit Hilfe von Iteratorklassen abstrahiert.
- Container unterscheiden sich hinsichtlich des Zugriffsmusters und der Komplexität der Operationen.

Folgende Container haben wir besprochen:

- Feld
- Liste (doppelt verkettet)
- Stack
- Queue
- Double-ended Queue
- Priority Queue
- Set
- Map

Fragentypen in der Klausur:

- Verständnisfragen
- Wenn freier Code, dann nur kleine Fragemente
- Fehler finden / Ergänzen von Code
- Verständnis von Code
- Sichtbarkeit, Defaultmethoden
- Konzepte der objektorientierten Programmierung
- Komplexität
- Algorithmen (insbesondere Stoff ab heute)

- keine explizite funktionale Programmierung

Anwendung: Huffman-Code

Problem: Wir wollen eine Zeichenfolge, z. B.

'ABRACADABRASIMSALABIM'

durch eine Folge von Zeichen aus der Menge $\{0, 1\}$ darstellen (kodieren).

Dazu wollen wir jedem der neun (verschiedenen) Zeichen aus der Eingabekette eine Folge von Bits zuzuordnen.

Am einfachsten ist es, einen Code fester Länge zu konstruieren. Mit n Bits können wir 2^n verschiedene Zeichen kodieren. Im obigem Fall genügen also vier Bit, um jedes der neun verschiedenen Zeichen in der Eingabekette zu kodieren, z. B.

A	0001	D	0100	M	0111
B	0010	I	0101	R	1000
C	0011	L	0110	S	1010

Die Zeichenkette wird dann kodiert als

$$\underbrace{0001}_{\text{A}}\underbrace{0010}_{\text{B}}\underbrace{1000}_{\text{R}}\ldots$$

Insgesamt benötigen wir $21 \cdot 4 = 84$ Bits (ohne die Übersetzungstabelle!).

Beobachtung: Kommen manche Zeichen häufiger vor als andere (wie etwa bei Texten in natürlichen Sprachen) so kann man Platz sparen, indem man Codes variabler Länge verwendet.

Beispiel: Morsecode.

Beispiel: Für unsere Beispielzeichenkette 'ABRACADABRASIMSALABIM' wäre folgender Code gut:

A	1	D	010	M	100
B	10	I	11	R	101
C	001	L	011	S	110

Damit kodieren wir unsere Beispielkette als

$$\underbrace{1}_{\mathsf{A}}\ \underbrace{10}_{\mathsf{B}}\ \underbrace{101}_{\mathsf{R}}\ \underbrace{1}_{\mathsf{A}}\ \underbrace{001}_{\mathsf{C}}\ldots$$

Schwierigkeit: Bei der Dekodierung könnte man diese Bitfolge auch interpretieren als

$$\underbrace{110}_{\mathsf{S}}\ \underbrace{101}_{\mathsf{R}}\ \underbrace{100}_{\mathsf{M}}\ldots$$

Abhilfe: Es gibt zwei Möglichkeiten das Problem zu umgehen:

1. Man führt zusätzliche Trennzeichen zwischen den Zeichen ein (etwa die Pause beim Morsecode).

2. Man sorgt dafür, dass kein Code für ein Zeichen der Anfang (Präfix) eines anderen Zeichens ist. Einen solchen Code nennt man Präfixcode.

Frage: Wie sieht der optimale Präfixcode für eine gegebene Zeichenfolge aus, d. h. ein Code der die gegebene Zeichenkette mit einer Bitfolge minimaler Länge kodiert.

Antwort: Huffmancodes! (Sie sind benannt nach ihrem Entdecker David Huffman[18], der auch die Optimalität dieser Codes gezeigt hat.)

Beispiel: Für unsere Beispiel-Zeichenkette ist ein solcher Huffmancode

A	11	D	10011	M	000
B	101	I	001	R	011
C	1000	L	10010	S	010

Die kodierte Nachricht lautet hier

$$111010111110001110011111010111101000100001011100101110100 1000$$

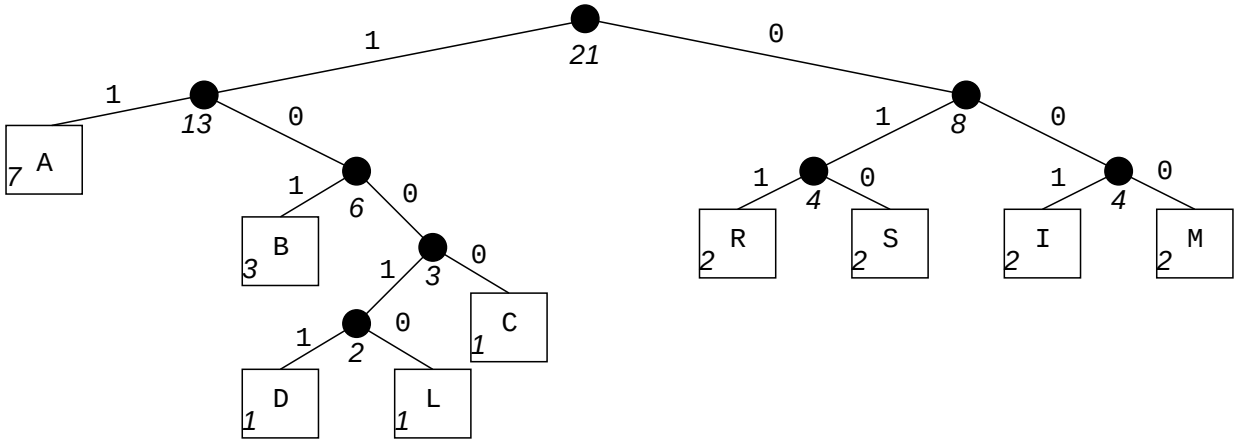
und hat nur noch 61 Bits!

[18] David A. Huffman, 1925–1999, US-amerk. Computerpionier.

Trie

Einem Präfixcode kann man einen binären Baum zuordnen, der Trie (von „retrieval“, Aussprache wie „try“) genannt wird. In den Blättern stehen die zu kodierenden Zeichen. Ein Pfad von der Wurzel zu einem Blatt kodiert das entsprechende Zeichen.

Blätter enthalten die zu kodierenden Zeichen, innere Knoten haben nur Wegweiserfunktion.



Bemerkung: Zeichen, die häufig vorkommen, stehen nahe bei der Wurzel. Zeichen, die seltener vorkommen, stehen tiefer im Baum.

Konstruktion von Huffmancodes

1. Zähle die Häufigkeit jedes Zeichens in der Eingabefolge. Erzeuge für jedes Zeichen einen Knoten mit seiner Häufigkeit. Packe alle Knoten in eine Menge E.

2. Solange die Menge E nicht leer ist: Entferne die zwei Knoten l und r mit **geringster** Häufigkeit aus E. Erzeuge einen neuen Knoten n mit l und r als Söhnen und der Summe der Häufigkeiten beider Söhne. Ist E leer, ist n die Wurzel des Huffmanbaumes, sonst stecke n in E.

Implementation

Programm: Huffman-Kodierung mit STL (HuffmanSTL.cc)

```
#include <iostream>
#include <map>
#include <queue>
#include <string>

using namespace std; // import namespace std

// There are no general binary trees in the STL.
// But we do not use much of this structure anyhow...
struct node
{
  struct node* left;
  struct node* right;
  char symbol;
  int weight;
```

```
  node( char c, int i ) {  // leaf constructor
    symbol = c;
    weight = i;
    left = right = 0;
  }
  node( node* l, node* r ) { // internal node constructor
    symbol = 0;
    weight = l->weight + r->weight;
    left = l;
    right = r;
  }
  bool isleaf() { return symbol != 0; }
  bool operator>( const node& a ) const {
    return weight > a.weight;
  }
};

// construct the Huffman trie for this message
node* huffman_trie( string message )
```

```
{
  // count multiplicities
  map<char, int> cmap;
  for ( string::iterator i=message.begin(); i!=message.end(); i++ )
    if ( cmap.find(*i) != cmap.end() )
      cmap[*i]++;
    else
      cmap[*i] = 1;

  // generate leaves with multiplicities
  priority_queue<node, vector<node>, greater<node> > q;
  for ( map<char, int>::iterator i=cmap.begin(); i!=cmap.end();
        i++ )
    q.push( node( i->first, i->second ) );

  // build Huffman tree (trie)
  while ( q.size() > 1 )
  {
    node* left = new node( q.top() );
```

```
    q.pop();
    node* right = new node( q.top() );
    q.pop();
    q.push( node( left, right ) );
  }
  return new node( q.top() );
}

// recursive filling of the encoding table 'code'
void fill_encoding_table( string s, node* i,
                          map<char, string>& code )
{
  if ( i->isleaf() )
    code[i->symbol] = s;
  else
  {
    fill_encoding_table( s + "0", i->left, code );
    fill_encoding_table( s + "1", i->right, code );
  }
```

```
}

// encoding
string encode( map<char, string> code, string& message ) {
  string encoded = "";
  for ( string::iterator i=message.begin(); i!=message.end(); i++ )
    encoded += code[*i];
  return encoded;
}

// decoding
string decode( node* trie, string& encoded ) {
  string decoded = "";
  node* node = trie;
  for ( string::iterator i=encoded.begin(); i!=encoded.end(); i++ )
  {
    if ( !node->isleaf() )
      node = (*i == '0') ? node->left : node->right;
    if ( node->isleaf() )
```

```
    {
      decoded.push_back( node->symbol );
      node = trie;
    }
  }
  return decoded;
}

int main() {
  string message = "ABRACADABRASIMSALABIM";

  // generate Huffman trie
  node* trie = huffman_trie( message );

  // generate and show encoding table
  map<char, string> table;
  fill_encoding_table( "", trie, table );
  for ( map<char, string>::iterator i=table.begin();
        i!=table.end(); i++ )
```

```
    cout << i->first << " " << i->second << endl;

  // encode and decode
  string encoded = encode( table, message );
  cout << "Encoded: " << encoded <<
    " [" << encoded.size() << " Bits]" << endl ;
  cout << "Decoded: " << decode( trie, encoded ) << endl;

  // the trie is not deleted here ...
}
```

Ausgabe: Wir erhalten einen anderen Huffman-Code als oben angegeben (der aber natürlich genauso effizient kodiert):

```
A 11
B 100
C 0010
D 1010
I 010
```

```
L 0011
M 1011
R 011
S 000
Encoded: 11100011110010111010111100011... [61 Bits]
Decoded: ABRACADABRASIMSALABIM
```

Effiziente Algorithmen und Datenstrukturen

Beobachtung: Einige der bisher vorgestellten Algorithmen hatten einen sehr hohen Aufwand (z. B. $O(n^2)$ bei Bubblesort, $O(n)$ bei Einfügen/Löschen aus der Priority-Queue). In vielen Fällen ist die STL-Implementation viel schneller.

Frage: Wie erreicht man diese Effizienz?

Ziel: In diesem Kapitel lernen wir Algorithmen und Datenstrukturen kennen, mit denen man hohe (in vielen Fällen sogar optimale) Effizienz erreichen kann.

Heap

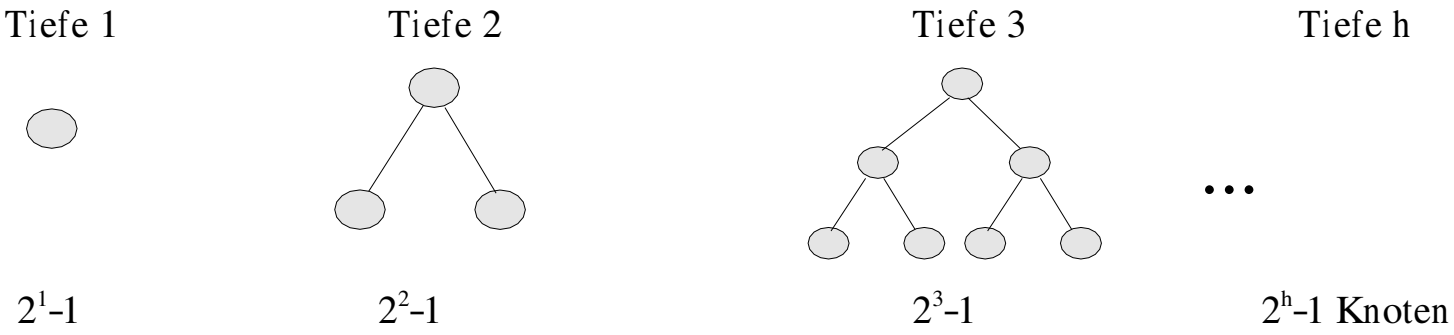
Die Datenstruktur Heap erlaubt es, Einfügen und Löschen in einer Prioritätswarteschlange mit $O(\log n)$ Operationen zu realisieren. Sie ist auch Grundlage eines schnellen Sortierverfahrens (Heapsort).

Definition: Ein Heap ist

- ein fast vollständiger binärer Baum
- Jedem Knoten ist ein Schlüssel zugeordnet. Auf der Menge der Schlüssel ist eine totale Ordnung (z. B. durch einen Operator $\leq$) definiert.
  Totale Ordnung: reflexiv ($a \leq a$), transitiv ($a \leq b, b \leq c \Rightarrow a \leq c$), total ($a \leq b \vee b \leq a$).
- Der Baum ist partiell geordnet, d. h. der Schlüssel jedes Knotens ist *nicht kleiner* als die Schlüssel in seinen Kindern (Heap-Eigenschaft).

Bezeichnung: Ein vollständiger binärer Baum ist ein

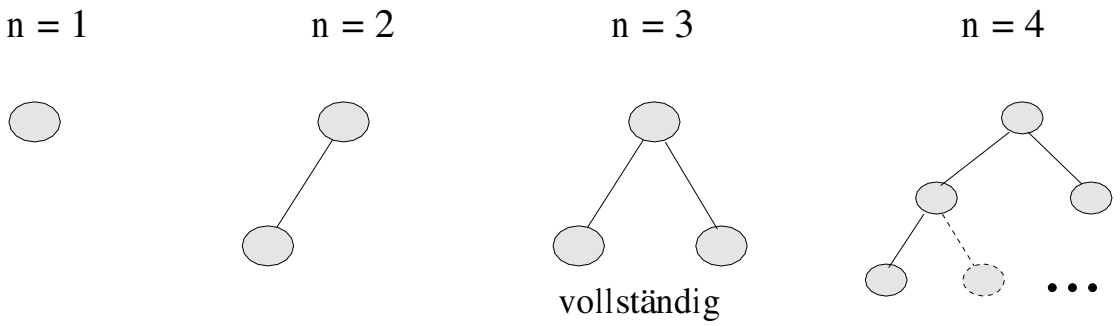
- binärer Baum der Tiefe h mit maximaler Knotenzahl,
- bei dem sich alle Blätter auf der gleichen Stufe befinden.



Bezeichnung: Ein fast vollständiger binärer Baum ist ein binärer Baum mit folgenden Eigenschaften:

- alle Blätter sind auf den beiden höchsten Stufen
- maximal ein innerer Knoten hat nur ein Kind
- Blätter werden von links nach rechts aufgefüllt.

Ein solcher Baum mit n Knoten hat eine eindeutige Struktur:

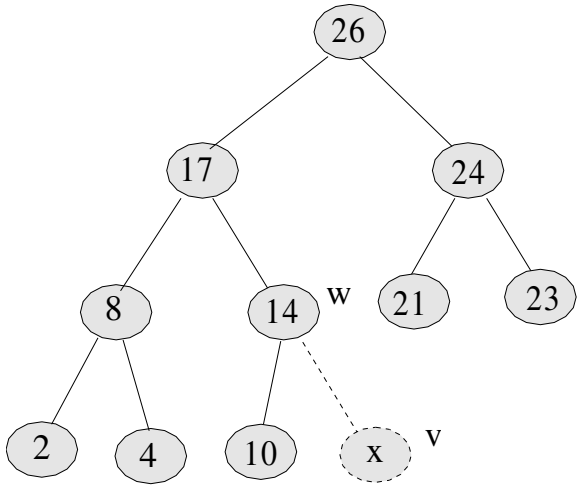


Einfügen

Problem: Gegeben ist ein Heap mit n Elementen und neues Element x. Konstruiere daraus einen um x erweiterten Heap mit $n+1$ Elementen.

Beobachtung: Die Struktur des Baumes mit $n+1$ Elementen liegt fest. Wenn man daher x an der neuen Position v einfügt, so kann die Heapeigenschaft nur im Knoten $w = \mathrm{Elter}(v)$ verletzt sein.

Beispiel:



Algorithmus: Wiederherstellen der Heapeigenschaft in maximal $\lceil \log_2 n \rceil$ Vertauschungen:

Falls $Inhalt(w) < Inhalt(v)$ dann
    tausche Inhalt
    Falls w nicht die Wurzel ist:
        setze $w = Elter(w); v = Elter(v);$
    sonst $\rightarrow$ fertig
sonst $\rightarrow$ fertig

Reheap

Die im folgenden beschriebene Reheap-Operation wird beim Entfernen der Wurzel gebraucht.

Problem: Gegeben ist ein fast vollständiger Baum mit Schlüsseln, so dass die Heapeigenschaft in allen Knoten exklusive der Wurzel gilt. Ziel ist die Transformation in einen echten Heap.

Algorithmus:

1. Tausche Schlüssel in der Wurzel mit dem größeren der beiden Kinder.

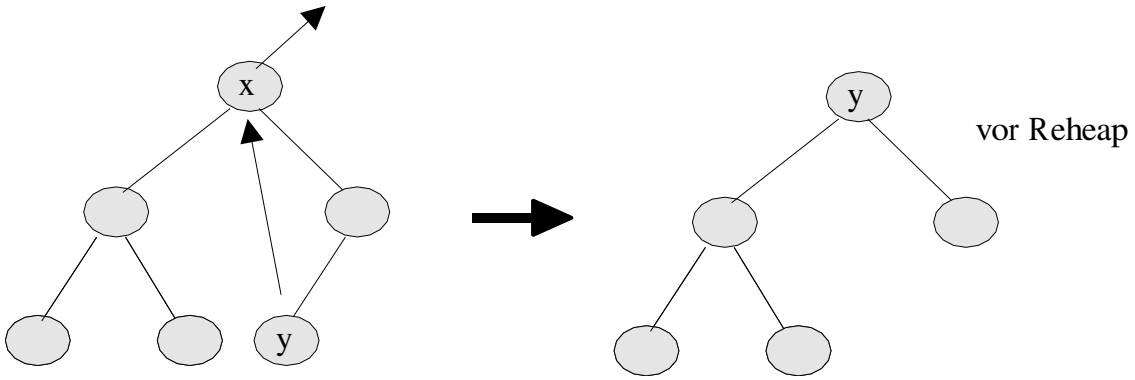
2. Wenn die Heap-Eigenschaft für dieses Kind nicht erfüllt ist, so wende den Algorithmus rekursiv an, bis ein Blatt erreicht wird.

Entfernen des Wurzelelements

Algorithmus:

- Ersetze den Wert in der Wurzel (Rückgabewert) durch das letzte Element des fast vollständigen binären Baumes.
- Verkleinere den Heap und rufe Reheap auf

Beispiel:



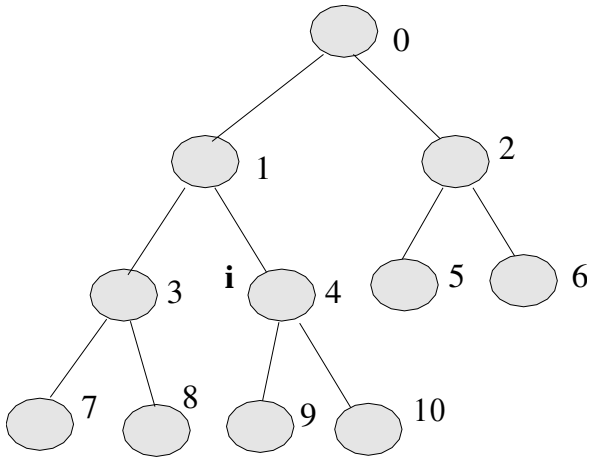
Komplexität

Ein fast vollständiger Baum mit n Knoten hat genau $\lceil \mathrm{ld}(n+1) \rceil$ Ebenen. Somit benötigt das Einfügen maximal $\lceil \mathrm{ld}(n+1) \rceil - 1$ Vergleiche und die Operation reheap maximal $2(\lceil \mathrm{ld}(n+1) \rceil - 1)$ Vergleiche.

Datenstruktur

Beobachtung: Die Knoten eines fast vollständigen binären Baumes können sehr effizient in einem Feld gespeichert werden:

Numeriert man die Knoten wie folgt:



Dann gilt für Knoten i:

Linkes Kind:	$2i + 1$
Rechtes Kind:	$2(i + 1)$
Elter:	$\lfloor \frac{i-1}{2} \rfloor$

Implementation

Programm: **Definition und Implementation (Heap.hh)**

```
template <class T>
class Heap
{
public:
  bool empty();
  void push( T x );
  void pop();
  T top();
private:
  std::vector<T> data;
  void reheap( int i );
};
```

```
template <class T>
void Heap<T>::push( T x )
{
  int i = data.size();
  data.push_back( x );
  while ( i > 0 && data[i] > data[(i-1)/2] )
  {
    std::swap( data[i], data[(i-1)/2] );
    i = (i - 1) / 2;
  }
}

template <class T>
void Heap<T>::reheap( int i )
{
  int n = data.size();
  while ( 2*i+1 < n )
```

```
    {
      int l = 2 * i + 1;
      int r = l + 1;
      int k = ( (r < n) && (data[r] > data[l]) ) ? r : l;
      if ( data[k] <= data[i] ) break;
      std::swap( data[k], data[i] );
      i = k;
    }
}

template <class T>
void Heap<T>::pop()
{
    std::swap( data.front(), data.back() );
    data.pop_back();
    reheap( 0 );
}
```

```
template <class T>
T Heap<T>::top()
{
  return data[0];
}

template <class T>
inline bool Heap<T>::empty()
{
  return data.size() == 0;
}
```

Programm: Anwendung (UseHeap.cc)

```
#include <vector>
#include <iostream>
```

```
#include "Heap.hh"
#include "Zufall.cc"

int main()
{
  Zufall z( 87123 );
  Heap<int> h;

  for ( int i=0; i<10; i=i+1 )
  {
    int k = z.ziehe_zahl();
    std::cout << k << std::endl;
    h.push( k );
  }
  std::cout << std::endl;
  while ( !h.empty() )
```

```
    {
      std::cout << h.top() << std::endl;
      h.pop();
    }
  }
```

Beobachtung: Mit Hilfe der Heap-Struktur lässt sich sehr einfach ein recht guter Sortieralgorithmus erzeugen. Dazu ordnet man Elemente einfach in einen Heap ein und extrahiert sie wieder. Dies wird später noch genauer beschrieben.

Sortierverfahren mit quadratischer Komplexität

Gegeben: Eine „Liste“ von Datensätzen $(D_0, \ldots, D_{n-1})$. Zu jedem Datensatz D_i gehört ein Schlüssel $k_i = k(D_i)$. Auf der Menge der Schlüssel sei eine $totale$ Ordnung durch einen Operator $\leq$ definiert.

Definition: Eine Permutation von $I = \{0, \ldots, n-1\}$ ist eine bijektive Abbildung $\pi : I \to I$.

Gesucht: Eine Permutation $\pi : \{0, \ldots, n-1\} \to \{0, \ldots, n-1\}$, so dass gilt

$$k_{\pi(0)} \leq \ldots \leq k_{\pi(n-1)}$$

Bemerkung: In der Praxis hat man:

- Die Datensätze sind normalerweise in einer Liste oder einem Feld gespeichert. Wir betrachten im folgenden den Fall des Felds.
- Oft braucht man die Permutation π nicht weiter, und es reicht aus, als Ergebnis einer Sortierfunktion eine sortierte Liste/Feld zu erhalten.
- Die Relation $\leq$ wird durch einen Vergleichsoperator definiert.
- Für große Datensätze sortiert man lieber ein Feld von Zeigern.
- Internes Sortieren: Alle Datensätze sind im Hauptspeicher.
- Externes Sortieren: Sortieren von Datensätzen, die auf Platten, Bändern, etc. gespeichert sind.

Die folgenden Implementierungen können z. B. mit einem `std::vector` aufgerufen werden.

Selectionsort (Sortieren durch Auswahl)

Idee:

- Gegeben sei ein Feld $a = (a_0, \ldots, a_{n-1})$ der Länge n.
- Suche das Minimum im Feld und tausche mit dem erstem Element.
- Danach steht die kleinste der Zahlen ganz links, und es bleibt noch ein Feld der Länge $n - 1$ zu sortieren.

Programm: **Selectionsort (Selectionsort.cc)**

```
template <class C>
void selectionsort( C& a )
{
  for ( int i=0; i<a.size()-1; i=i+1 )
```

```
    {    // i Elemente sind sortiert
      int min = i;
      for ( int j=i+1; j<a.size(); j=j+1 )
        if ( a[j] < a[min] ) min = j;
      std::swap( a[i], a[min] );
    }
  }
```

Bemerkung:

- Komplexität: in führender Ordnung $\frac{n^2}{2}$ Vergleiche, n Vertauschungen $\rightarrow O(n^2)$.
- Die Anzahl von Datenbewegungen ist optimal, das Verfahren ist also zu empfehlen, wenn möglichst wenige Datensätze bewegt werden sollen.

Bubblesort

Idee: (Siehe den Abschnitt über Effizienz generischer Programmierung.)

- Gegeben sei ein Feld $a = (a_0, \ldots, a_{n-1})$ der Länge n.
- Durchlaufe die Indizes $i = 0, 1, \ldots, n-2$ und vergleiche jeweils a_i und a_{i+1}. Ist $a_i > a_{i+1}$ so vertausche die beiden.
- Nach einem solchen Durchlauf steht die größte der Zahlen ganz rechts, und es bleibt noch ein Feld der Länge $n-1$ zu sortieren.

Programm: **Bubblesort mit STL (Bubblesort.cc)**

```
template <class C>
void bubblesort( C& a )
{
  for ( int i=a.size()-1; i>=0; i-- )
    for ( int j=0; j<i; j=j+1 )
      if ( a[j+1] < a[j] )
        std::swap( a[j+1], a[j] );
}
```

Bemerkung:

- Komplexität: in führender Ordnung $\frac{n^2}{2}$ Vergleiche, $\frac{n^2}{2}$ Vertauschungen

Insertionsort (Sortieren durch Einfügen)

Beschreibung: Der bereits sortierte Bereich liegt links im Feld und das nächste Element wird jeweils soweit nach links bewegt, bis es an der richtigen Stelle sitzt.

Programm: **Insertionsort mit STL (Insertionsort.cc)**

```
template <class C>
void insertionsort( C& a )
{
  for ( int i=1; i<a.size(); i=i+1 )
  {
    // i Elemente sind sortiert
    int j = i;
    while ( j > 0 && a[j-1] > a[j] )
    {
      std::swap( a[j], a[j-1] );
      j = j - 1;
```

```
        }
    }
}
```

Bemerkung:

- Komplexität: $\frac{n^2}{2}$ Vergleiche, $\frac{n^2}{2}$ Vertauschungen $\rightarrow O(n^2)$.

- Ist das Feld bereits sortiert, endet der Algorithmus nach $O(n)$ Vergleichen. Sind in ein bereits sortiertes Feld mit n Elementen m weitere Elemente einzufügen, so gelingt dies mit Insertionsort in $O(nm)$ Operationen. Dies ist optimal für sehr kleines m ($m \ll \log n$).

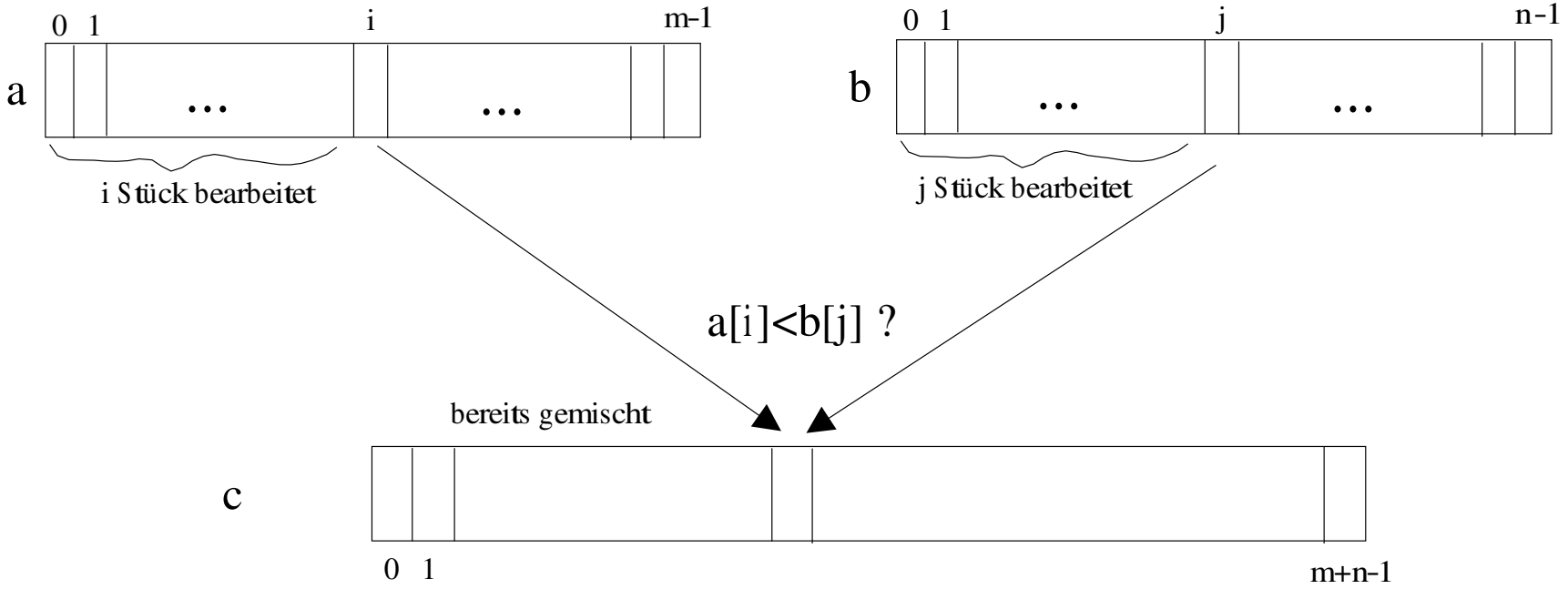
Sortierverfahren optimaler Ordnung

In diesem Abschnitt betrachten wir Sortierverfahren, die den Aufwand $O(n \log n)$ haben. Man kann zeigen, dass dieser Aufwand für allgemeine Felder von Datensätzen optimal ist.

Erinnerung: $\log x = \log_e x$, $\mathrm{ld}\, x = \log_2 x$. Wegen $\mathrm{ld}\, x = \frac{\log x}{\log 2} = 1.44\ldots \cdot \log x$ gilt $O(\log n) = O(\mathrm{ld}\, n)$.

Mergesort (Sortieren durch Mischen)

Beobachtung: Zwei bereits sortierte Felder der Länge m bzw. n können sehr leicht (mit Aufwand $O(m+n)$) zu einem sortierten Feld der Länge $m+n$ „vermischt“ werden:



Dies führt zu folgendem Algorithmus vom Typ „Divide and Conquer“:

Algorithmus:

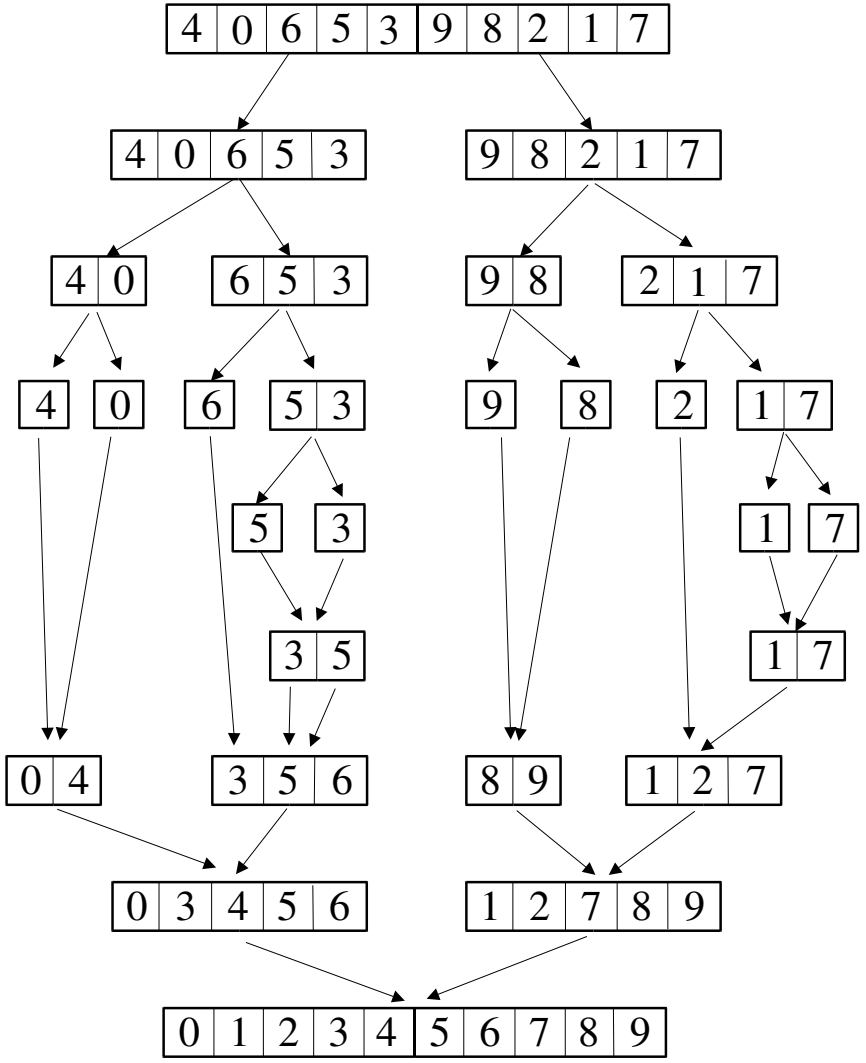
- Gegeben ein Feld a der Länge n.
- Ist $n = 1$, so ist nichts zu tun, sonst
- Zerlege a in zwei Felder $a1$ mit Länge $n1 = n/2$ (ganzzahlige Division) und $a2$ mit Länge $n2 = n - n1$,
- sortiere $a1$ und $a2$ (Rekursion) und
- mische $a1$ und $a2$.

Programm: Mergesort mit STL (Mergesort.cc)

```
template <class C>
void rec_merge_sort( C& a, int o, int n )
{   // sortiere Eintraege [ o, o+n-1 ]
  if ( n == 1 ) return;

  // teile und sortiere rekursiv
  int n1 = n / 2;
  int n2 = n - n1;
  rec_merge_sort( a, o, n1 );
  rec_merge_sort( a, o + n1, n2 );

  // zusammenfuegen
  C b( n ); // Hilfsfeld
  int i1 = o, i2 = o + n1;
  for ( int k=0; k<n; k=k+1 )
    if ( ( i2 >= o+n ) || ( i1 < o+n1 && a[i1] <= a[i2] ) )
      b[k] = a[i1++];
```

```
      else
        b[k] = a[i2++];

    // umkopieren
    for ( int k=0; k<n; k=k+1 ) a[o + k] = b[k];
}

template <class C>
void mergesort( C& a )
{
    rec_merge_sort( a, 0, a.size() );
}
```

Beispiel:


4 0 6 5 3 9 8 2 1 7
4 0 6 5 3
9 8 2 1 7
4 0
6 5 3
9 8
2 1 7
4
0
6
5 3
9
8
2
1 7
5
3
1
7
3 5
1 7
0 4
3 5 6
8 9
1 2 7
0 3 4 5 6
1 2 7 8 9
0 1 2 3 4 5 6 7 8 9

Bemerkung:

- Mergesort benötigt zusätzlichen Speicher von der Größe des zu sortierenden Felds.

- Die Zahl der Vergleiche ist aber (in führender Ordnung) $n \operatorname{ld} n$.
  Beweis für $n = 2^k$: Für die Zahl der Vergleiche $V(n)$ gilt (Induktion)

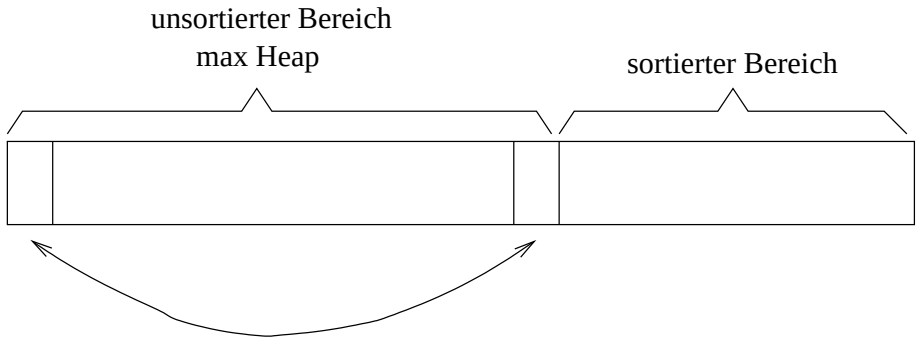
$$\begin{aligned} V(1) &= 0 = 1 \operatorname{ld} 1 \\ V(n) &= 2V(\frac{n}{2}) + n - 1 \leq 2(k-1)\frac{n}{2} + n \leq nk \end{aligned}$$

  ($2V(n/2)$: Sortieren beider Hälften; $n-1$: Merge). Man kann zeigen, dass dies optimal ist.

- Mergesort ist stabil, d. h. Datensätze mit gleichen Schlüsseln verbleiben in derselben Reihenfolge wie zuvor

Heapsort

Idee: Transformiere das Feld in einen Heap, und dann wieder in ein sortiertes Feld. Wegen der kompakten Speicherweise für den Heap kann dies ohne zusätzlichen Speicherbedarf geschehen, indem man das Feld wie folgt unterteilt:



Bemerkung:

- Die Transformation des Felds in einen Heap kann auf zwei Weisen geschehen:

  1. Der Heap wird von vorne durch `push` aufgebaut.
  2. Der Heap wird von hinten durch `reheap` aufgebaut.

  Da wir `reheap` sowieso für die pop-Operation brauchen, wählen wir die zweite Variante.

- Heapsort hat in führender Ordnung die Komplexität von $2n \operatorname{ld} n$ Vergleichen. Der zusätzliche Speicheraufwand ist unabhängig von n (in-situ-Verfahren)!

- Im Gegensatz zu Mergesort ist Heapsort nicht stabil.

Programm: Heapsort mit STL (Heapsort.cc)

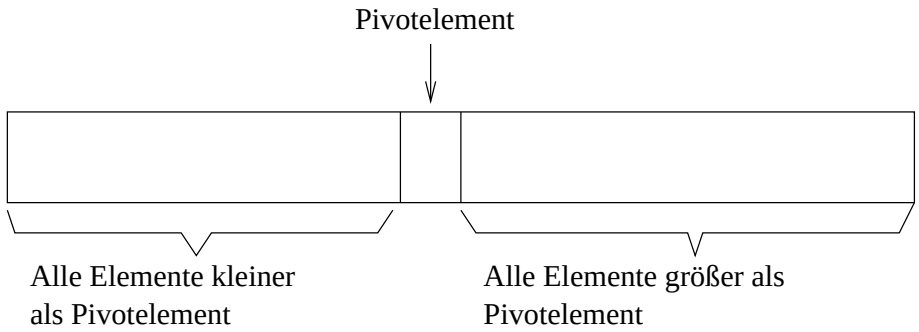
```
template <class C>
inline void reheap( C& a, int n, int i )
{
  while ( 2*i+1 < n )
  {
    int l = 2 * i + 1;
    int r = l + 1;
    int k = ( ( r < n ) && ( a[r] > a[l] ) ) ? r : l;
    if ( a[k] <= a[i] ) break;
    std::swap( a[k], a[i] );
    i = k;
  }
}

template <class C>
```

```
void heapsort( C& a )
{
  // build the heap by reheaping from the rear
  for ( int i=a.size()-1; i>=0; i-- )
    reheap( a, a.size(), i );
  // build the sorted list by popping the heap
  for ( int i=a.size()-1; i>=0; i-- )
  {
    std::swap( a[0], a[i] );
    reheap( a, i, 0 );
  }
}
```

Quicksort

Beobachtung: Das Hauptproblem bei Mergesort war das speicheraufwendige Mischen. Man könnte es vermeiden, wenn man das Feld so in zwei (möglichst gleichgroße) Teile unterteilen könnte, das alle Elemente des linken Teilfelds kleiner oder gleich allen Elementen des rechten Teilfeldes sind.

Idee: Wähle „zufällig“ ein beliebiges Element $q \in \{0, \ldots, n-1\}$ aus, setze Pivot$= a[q]$ und zerlege das Feld so, dass das Eingabefeld folgende Gestalt hat:



(Elemente gleich $a[q]$ dürfen in linkes oder rechtes Teilfeld eingefügt werden).

Programm: Quicksort mit STL (Quicksort.cc)

```
template <class C>
int qs_partition( C& a, int l, int r, int q )
{
  std::swap( a[q], a[r] );
  q = r;        // Pivot ist jetzt ganz rechts
  int i = l - 1, j = r;

  while ( i < j )
  {
    i = i + 1; while ( i < j && a[i] <= a[q] ) i = i + 1;
    j = j - 1; while ( i < j && a[j] >= a[q] ) j = j - 1;
    if ( i < j )
      std::swap( a[i], a[j] );
    else
      std::swap( a[i], a[q] );
  }
  return i; // endgueltige Position des Pivot
```

```
}

template <class C>
void qs_rec( C& a, int l, int r )
{
  if ( l < r )
  {
    int i = qs_partition( a, l, r, r );
    qs_rec( a, l, i-1 );
    qs_rec( a, i+1, r );
  }
}

template <class C>
void quicksort( C& a ) {
  qs_rec( a, 0, a.size()-1 );
}
```

Bemerkung:

- Man kann im allgemeinen $nicht$ garantieren, dass beide Hälften gleich groß sind.
- Im schlimmsten Fall wird das Pivotelement immer so gewählt, dass man ein einelementiges Teilfeld und den Rest als Zerlegung erhält. Dann hat Quicksort den Aufwand $O(n^2)$.
- Im besten Fall ist die Zahl der Vergleiche so gut wie Mergesort.
- Im „Mittel“ erhält man in führender Ordnung $1.386\, n \operatorname{ld} n$ Vergleiche.
- Auch Quicksort ist nicht stabil.
- Praktisch wählt man oft drei Elemente zufällig aus und wählt das mittlere. Damit wird Quicksort ein randomisierter Algorithmus.

- Die Wahrscheinlichkeit den $O(n^2)$ Fall zu erhalten ist bei zufälliger Pivotwahl $1/n!$.

Anwendung

Mit folgendem Programm kann man die verschiedenen Sortierverfahren ausprobieren:

Programm: **UseSort.cc**

```
#include <iostream>
#include <vector>
#include "Bubblesort.cc"
#include "Selectionsort.cc"
#include "Insertionsort.cc"
#include "Mergesort.cc"
#include "Heapsort.cc"
#include "Quicksort.cc"
#include "Zufall.cc"
#include "timestamp.cc"

void initialize( std::vector<int>& a )
{
```

```
  Zufall z( 8267 );
  for ( int i=0; i<a.size(); ++i )
    a[i] = z.ziehe_zahl();
}

int main()
{
  int n = 100000;
  std::vector<int> a( n );

  initialize( a );
  time_stamp();
  quicksort( a );
  std::cout << "n=" << n << " quicksort t="
            << time_stamp() << std::endl;

  initialize( a );
  time_stamp();
  mergesort( a );
```

```
std::cout << "n=" << n << " mergesort t="
          << time_stamp() << std::endl;

initialize( a );
time_stamp();
heapsort( a );
std::cout << "n=" << n << " heapsort t="
          << time_stamp() << std::endl;

initialize( a );
time_stamp();
bubblesort( a );
std::cout << "n=" << n << " bubblesort t="
          << time_stamp() << std::endl;

initialize( a );
time_stamp();
insertionsort( a );
std::cout << "n=" << n << " insertionsort t="
```

```
            << time_stamp() << std::endl;

  initialize( a );
  time_stamp();
  selectionsort( a );
  std::cout << "n=" << n << " selectionsort t="
            << time_stamp() << std::endl;
}
```

Suchen

Binäre Suche in einem Feld

Idee: In einem sortierten Feld kann man Elemente durch sukzessives Halbieren schnell finden.

Bemerkung:

- Aufwand: in jedem Schritt wird die Länge des Suchbereichs halbiert. Der Aufwand beträgt daher $\lceil \mathrm{ld}(n) \rceil$ Vergleiche, denn dann kann man nicht weiter halbieren. Anschließend braucht man noch einen Vergleich, um zu prüfen, ob das Element das Gesuchte ist. $\Rightarrow \lceil \mathrm{ld}(n) \rceil + 1$ Vergleiche.

- Die binäre Suche ermöglicht also auch die Aussage, dass ein Element nicht enthalten ist!

Programm: Nicht-rekursive Formulierung (Binsearch.cc)

```
template <class C>
int binsearch( C& a, typename C::value_type x )
{   // returns either index (if found) or -1
  int l = 0;
  int r = a.size();
  while ( 1 )
  {
    int m = ( l + r ) / 2;
    if ( m == l )
      return ( a[m] == x ) ? m : -1;
    if ( x < a[m] )
      r = m;
    else
      l = m;
  }
```

$\}$

Bemerkung:

- Die binäre Suche beschleunigt nur das Finden.
- Einfügen und Löschen haben weiterhin Aufwand $O(n)$, da Feldelemente verschoben werden müssen.
- Binäre Suche geht auch nicht mit einer Liste (kein wahlfreier Zugriff).

Binäre Suchbäume

Beobachtung: Die binäre Suche im Feld kann als Suche in einem binären Suchbaum interpretiert werden (der aber in einem Feld abgespeichert wurde).

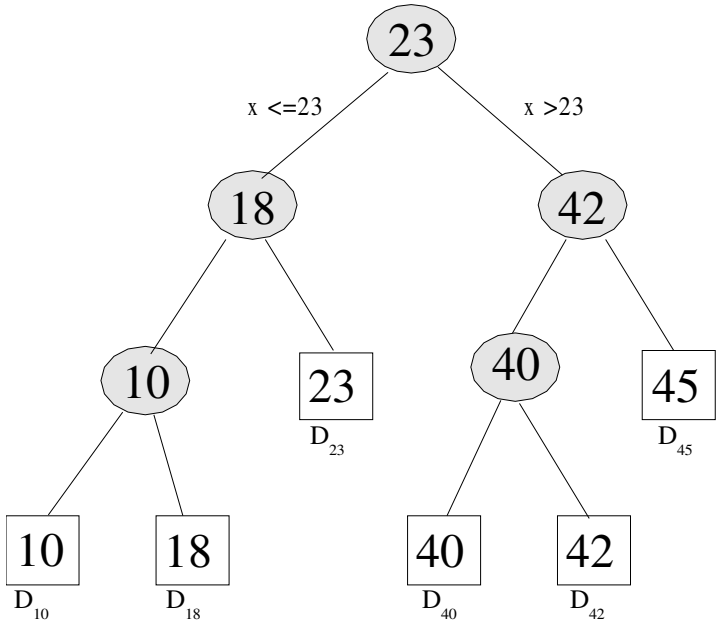
Idee: Die Verwendung einer echten Baum-Datenstruktur kann schnelles Einfügen und Entfernen von Knoten ermöglichen.

Definition: Ein binärer Suchbaum ist ein binärer Baum, in dessen Knoten Schlüssel abgespeichert sind und für den die Suchbaumeigenschaft gilt:

> Der Schlüssel in jedem Knoten ist größer gleich allen Schlüsseln im linken Teilbaum und kleiner als alle Schlüssel im rechten Teilbaum (Variante 1).

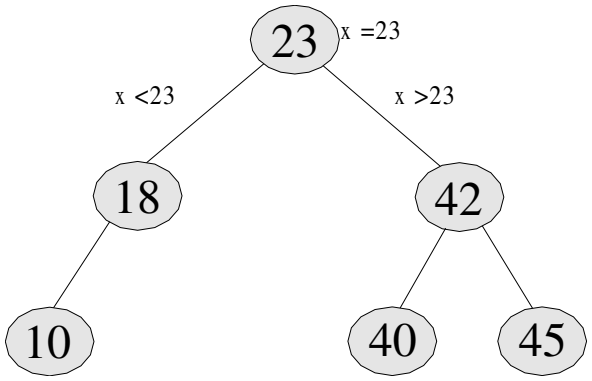
Bemerkung: Der Aufwand einer Suche entspricht der Höhe des Baumes.

Variante 1



- *Innere* Knoten enthalten nur Schlüssel
- (Verweise auf) Datensätze sind in den *Blättern* des Baumes gespeichert.

Variante 2 Man speichert Schlüssel und Datensatz genau einmal in einem Knoten:



Bemerkung: Innere Knoten unterscheiden sich von Blättern dann nur noch durch das Vorhandensein von Kindern.

Die Suchalgorithmen unterscheiden sich leicht:

Variante 1: Am Blatt prüfe Schlüssel; innerer Knoten: bei $\leq$ gehe links, sonst rechts. Innere Knoten haben immer zwei Kinder!

Variante 2: Prüfe Schlüssel; Falls Kind links existiert und Schlüssel $<$ gehe links; falls Kind rechts existiert und Schlüssel $>$ gehe rechts; sonst nicht gefunden.

Einfügen

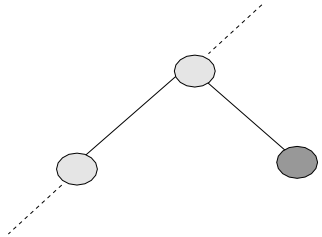
Das Einfügen (in Variante 2) geschieht, indem man durch den Baum bis zu einem Knoten läuft, wo man den Datensatz/Schlüssel einfügen kann.

Beispiel: Einfügen von 20 im Baum auf der letzten Folie. Diese wird rechts unter der 18 eingefügt. Wo wird die 41 eingefügt?

Das Löschen ist etwas komplizierter, weil verschiedene Situationen unterschiedlich behandelt werden müssen.

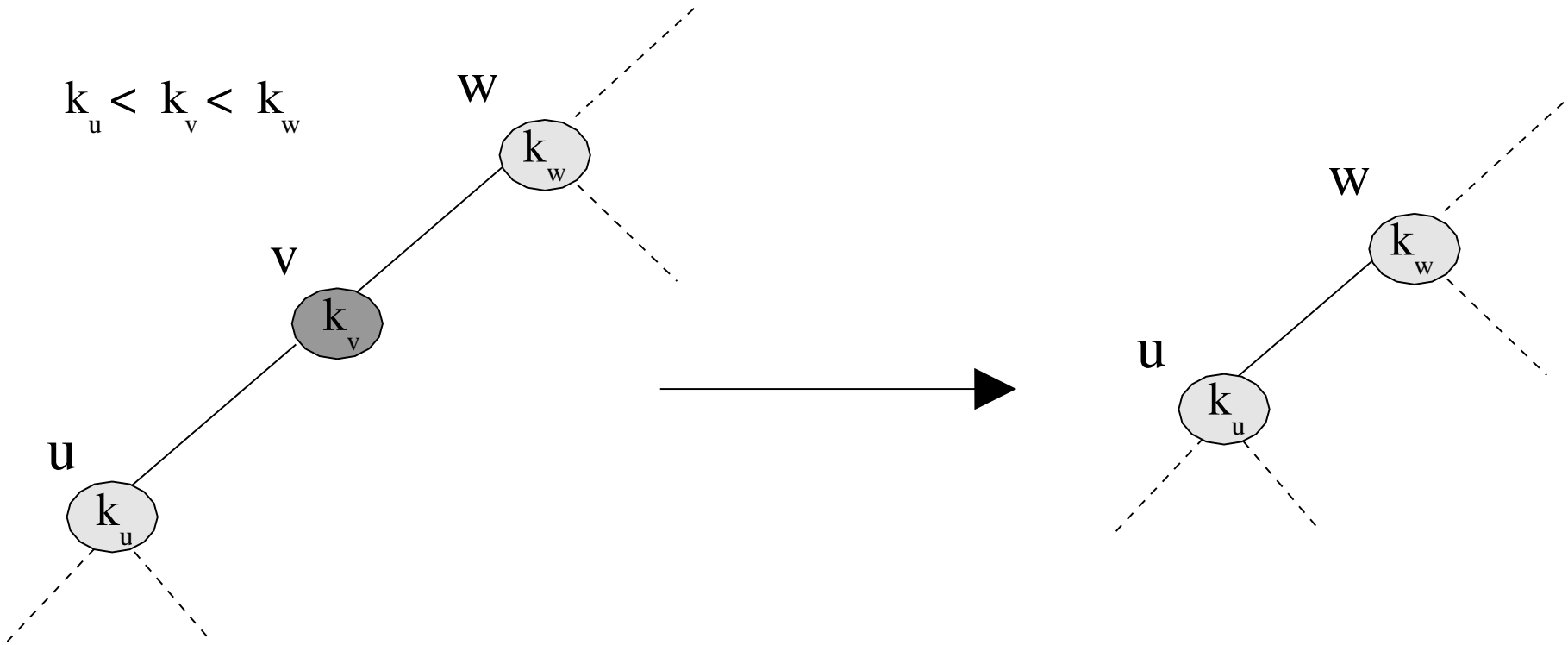
Löschen eines Blattes (Variante 2)

$\rightarrow$ einfach wegnehmen



Löschen eines Knotens mit einem Kind (Variante 2)

Der Knoten kann einfach herausgenommen werden.



Löschen eines Knotens mit zwei Kindern (Variante 2)

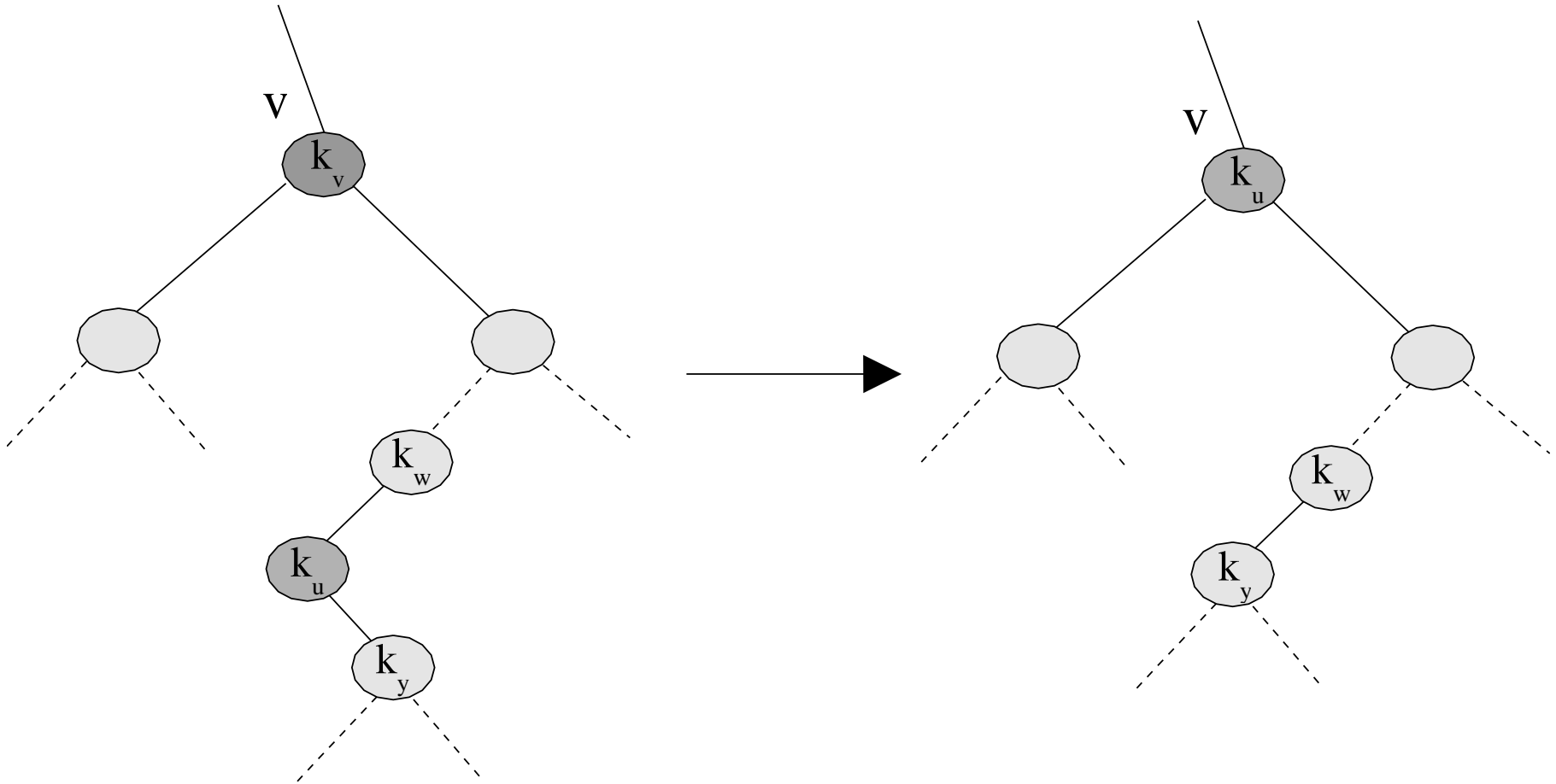
Schlüssel k_v soll gelöscht werden. Betrachte k_u: kleinster Schlüssel im rechten Teilbaum von v.

Behauptung: u hat höchstens einen rechten Teilbaum (also keine zwei Kinder)! Das ist klar, denn hätte u einen linken Teilbaum so wären die Schlüssel dort kleiner als k_u und somit wäre k_u nicht der minimale Schlüssel rechts von v.

Vorgehensweise:

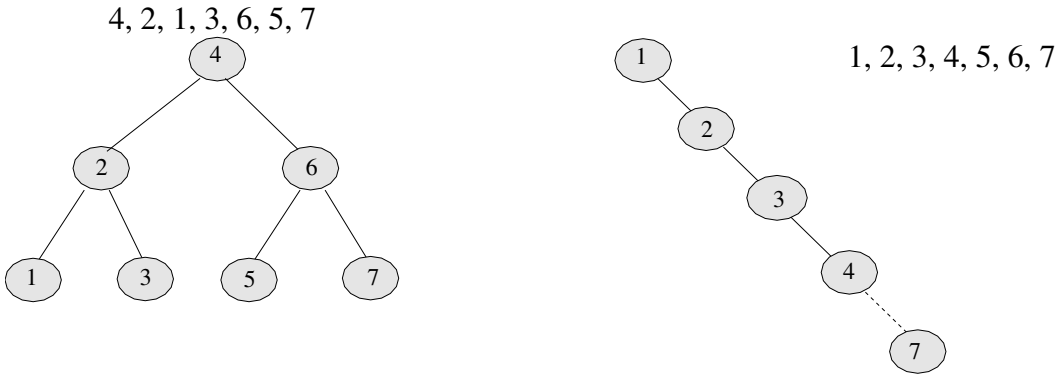
- ersetze k_v durch k_u und lösche Knoten u (dies ist einfach, da der keine zwei Kinder hat).

- Suchbaumeigenschaft in v bleibt erhalten: Alle rechts verbleibenden Knoten haben Schlüssel größer als k_u, da k_u minimal war. Alle links verbleibenden Schlüssel waren ohnehin kleiner.


v
k v
k w
k u
k y
v
k u
k w
k y

Problem: Die Gestalt des Suchbaumes und damit der Aufwand für seine Bearbeitung hängt entscheidend von der Reihenfolge des Einfügens (und eventuell Löschens) der Schlüssel ab!

Beispiel:



Der rechte Binärbaum entspricht im wesentlichen einer Listenstruktur! Der Aufwand zur Suche ist entsprechend $O(n)$.

Ausgeglichene Bäume

Beobachtung: Um optimale Effizienz zu gewährleisten, müssen sowohl Einfüge- als auch Löschoperation im Suchbaum sicherstellen, dass für die Höhe $H(n) = O(\log n)$ gilt. Dies kann auf verschiedene Weisen erreicht werden.

AVL-Bäume

Die AVL-Bäume wurden 1962 von Adelson-Velskii-Landis eingeführt. Es sind Binärbäume, die garantieren, dass sich die Höhen von rechtem und linken Teilbaum ihrer Knoten maximal um 1 unterscheiden (Höhenbalancierung).

(2,3)-Baum

Der (2,3)-Baum wurde 1970 von Hopcroft eingeführt. Er ist ein Spezialfall des später besprochenen (a,b)-Baums. Die Idee ist, mehr Schlüssel/Kinder pro Knoten zuzulassen.

B-Bäume

B-Bäume wurden 1970 von Bayer und McCreight eingeführt. Der B-Baum der

Ordnung m ist gleich dem (a, b)-Baum mit $a = \lceil \frac{m}{2} \rceil$, $b = m$. Hier haben Knoten bis zu m Kinder. Für großes m führt das zu sehr flachen Bäumen. Der B-Baum wird oft zur Suche in externem Speicher verwendet. Dabei ist m die Anzahl der Schlüssel, die in einen Sektor der Platte passen (z. B. Sektorgröße 512 Byte, Schlüssel 4 Byte $\Rightarrow$ m=128).

α-balancierter Baum

α-balancierte Bäume wurden um 1973 von Nievergelt und Reingold eingeführt. Idee: Die Größe $|T(v)|$ des Baums am Knoten v und des rechten (oder linken) Teilbaums $|T_l(v)|$ erfüllen

$$\alpha \leq \frac{|T_l(v)| + 1}{|T(v)| + 1} \leq 1 - \alpha$$

Dies garantiert wieder $H = O(\log n)$.

Rot-Schwarz-Bäume

Rot-Schwarz-Bäume wurden 1978 von Bayer, Guibas, Sedgewick eingeführt. Hier

haben Knoten verschiedene „Farben“. Die Einfüge- und Löschoperationen erhalten dann gewisse Anforderungen an die Farbreihenfolge, welche wieder $H = O(\log n)$ garantieren. Man kann auch eine Äquivalenz zum (2,4)-Baum zeigen.

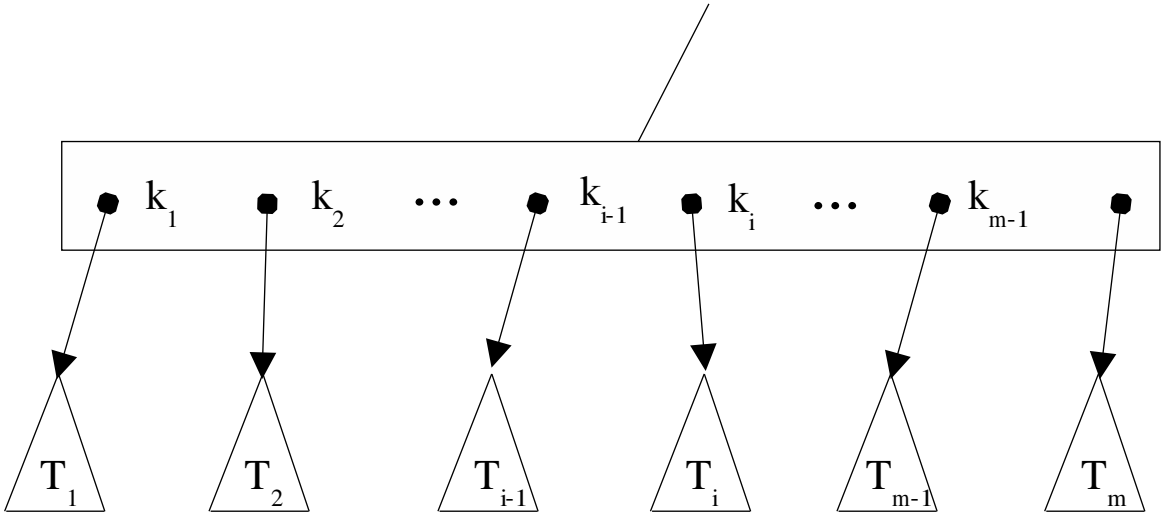
(a,b)-Bäume

(a,b)-Bäume wurden 1982 von Huddleston und Mehlhorn als Verallgemeinerung der B-Bäume und des (2,3)-Baums eingeführt. Alle inneren Knoten haben hier mindestens a und höchstens b Kinder für $a \geq 2$ und $b \geq 2a - 1$. Außerdem befinden sich alle Blätter auf der gleichen Stufe. Dies garantiert dann wieder $H = O(\log n)$.

(a, b)-Bäume (Mehlhorn 1982)

Der (a, b)-Baum stellt eine Erweiterung des binären Baumes auf viele Schlüssel pro Knoten dar:

- Jeder innere Knoten enthält Schlüssel
- Blätter enthalten Datensätze (Variante 1 oben)



Allgemeiner innerer Knoten

- höherer Verzweigungsgrad: Ein Knoten hat bis zu m Kinder und $m-1$ Schlüssel
- alle Schlüssel sind sortiert: $k_1 < k_2 < k_3 < \cdots < k_{m-1}$
- für alle Schlüssel k im Teilbaum T_i, $1 \leq i \leq m$, gilt: $k_{i-1} < k \leq k_i$ (setze $k_0 = -\infty, k_m = \infty$)

Ein Baum aus solchen Knoten heißt (a, b)-Baum, falls gilt:

- alle Blätter sind auf derselben Stufe
- jeder innere Knoten hat maximal b Kinder und für die minimale Zahl der Kinder innerer Knoten gilt
  - Wurzel hat mindestens 2 Kinder
  - andere haben $mindestens\ a \geq 2$ Kinder
- Es gilt: $b \geq 2a - 1$

z. B. $a = 2, b = 3$ minimaler (a, b)-Baum (Hopcroft 1970)

$a = 2, b = 4$ $(2, 4)$-Baum$\Longleftrightarrow$rot-schwarz-Baum

$a = \lceil \frac{m}{2} \rceil, b = m$ B-Baum, zur Suche in externem Speicher

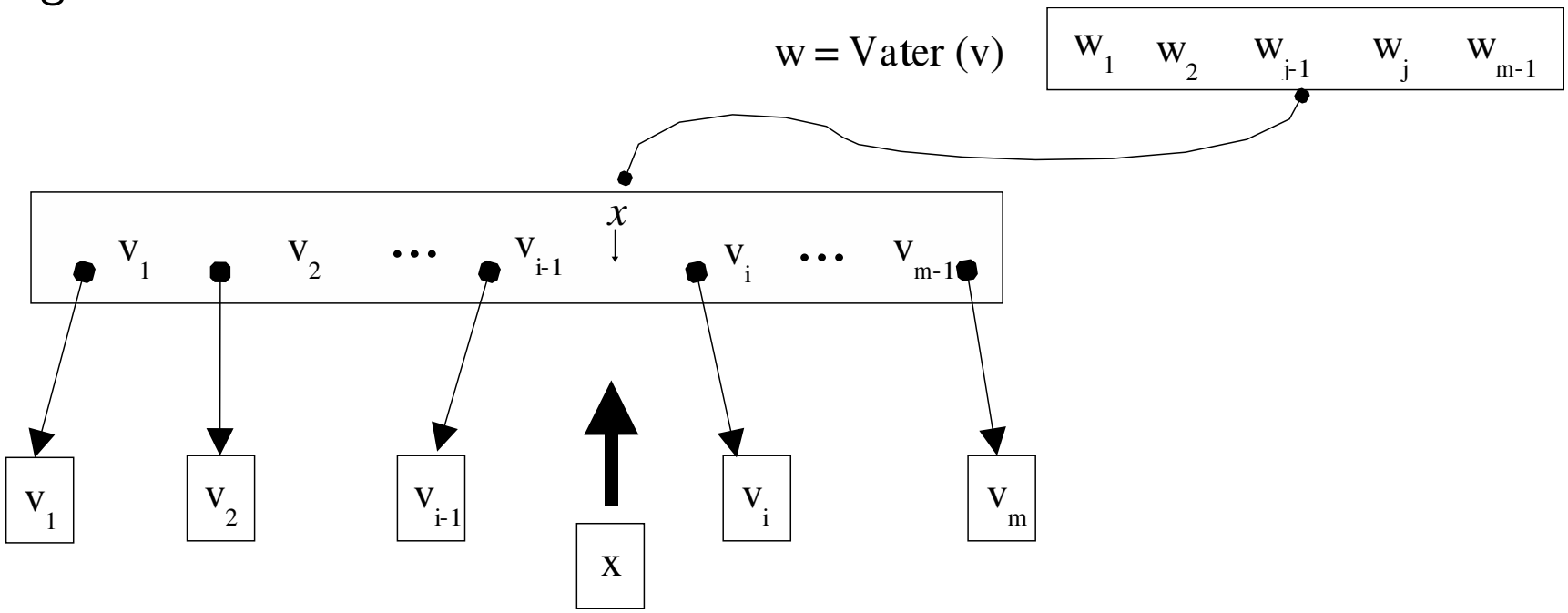
$\Rightarrow m$ Anzahl der Schlüssel, die in einen Sektor der Platte pass

z. B. Sektorgröße 512 Byte, Schlüssel 4 Byte $\Rightarrow$ b=128

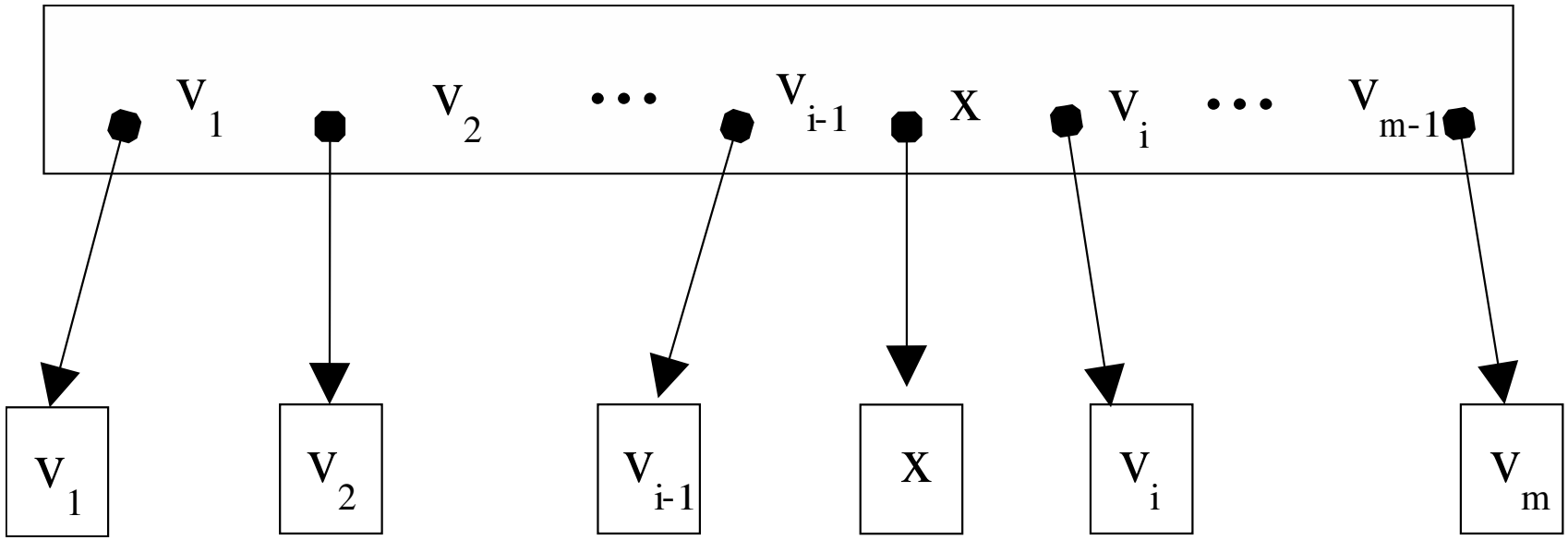
$\Rightarrow$sehr breite, flache Bäume

Einfügen in (a,b)-Baum

Schritt 1: Suche Schlüssel x. Falls x noch nicht drin ist, wird die Einfügeposition für x gefunden:



Schritt 2: Füge x in v ein. v ist ein Knoten, dessen Kinder Blätter sind.

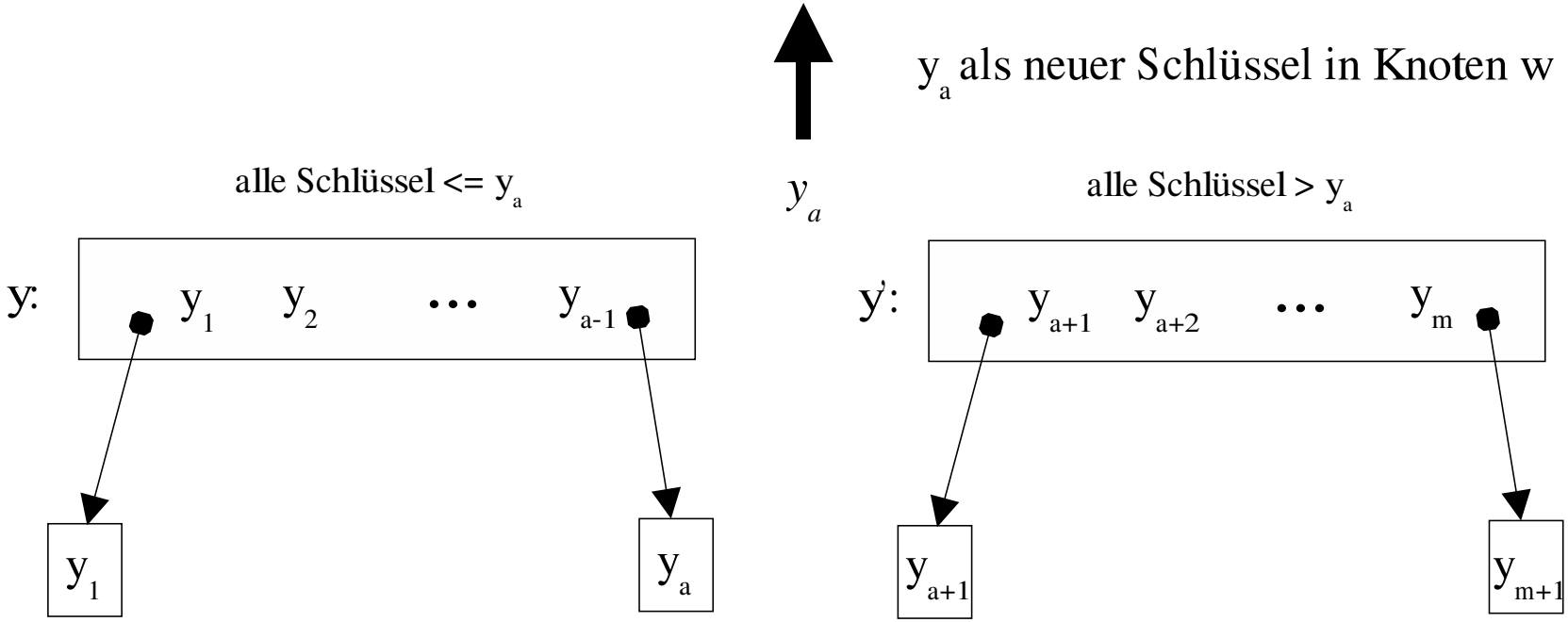


Es ergeben sich nun zwei Fälle:

- $m+1 \leq b \Rightarrow$ fertig!
- $m+1 = b+1$ (d. h. vorher war $m = b$) $\Rightarrow$ Schritt 3

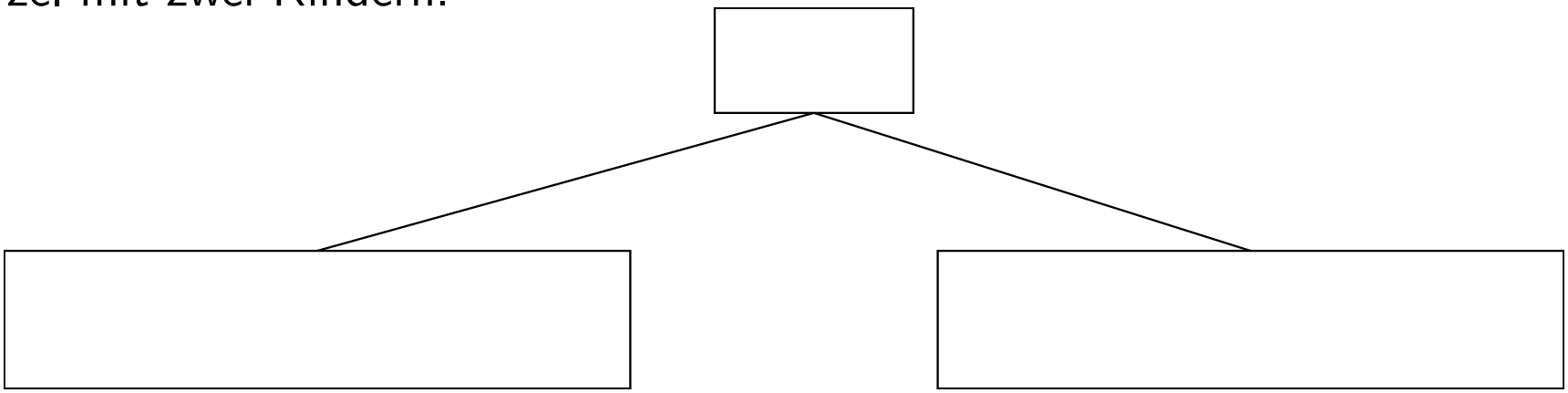
Schritt 3: Es war $m = b$ in v. Nenne den neu entstandenen Knoten y. Dieser hat entsprechend zunächst $b+1$ Kinder, das ist verboten.

- Spalte y in zwei Knoten: y, y'
- y behält a Kinder
- y' erhält $b + 1 - a \geq (2a - 1) + 1 - a = a \Rightarrow$ d. h. y' hat min. a Kinder



Schritt 4: Fahre rekursiv bis zur Wurzel fort (falls nötig)

Schritt 5: Nach (eventuell notwendigem) Splitten der Wurzel erhält man eine neue Wurzel mit zwei Kindern.



Bemerkung: Das Löschen eines Schlüssels verläuft analog in umgekehrter Reihenfolge.

Implementation von (a,b)-Bäumen

Als Beispiel geben wir mit dem folgenden Programm eine mögliche Implementation von (a,b)-Bäumen an.

Programm: (ab-tree.cc)

```
#include <set>
#include <iostream>
#include <vector>
#include <cassert>
using namespace std;

const int m = 2;        // B-tree of order m
const int a = m;        // minimal number of keys
const int b = 2 * m;    // maximal number of keys

template <class T>
struct Node
```

```
{
  // data
  vector<T> keys;
  vector<Node *> children;
  Node* parent;
  // interface
  Node( Node* p ) { parent = p; }
  bool is_leaf() { return children.size() == 0; }
  Node* root() { return (parent == 0) ? this : parent->root(); }
  Node* find_node( T item );
  int find_pos( T item );
  bool equals_item( int pos, T item );
};

// finds first position i such that keys[i] >= item
template <class T>
int Node<T>::find_pos( T item )
{
  int i = 0;
```

```
  while ( ( i < keys.size() ) && ( keys[i] < item ) ) i++;
  return i;
}

// checks if the key at position pos contains item
template <class T>
bool Node<T>::equals_item( int pos, T item )
{
  return ( pos < keys.size() ) && !( item < keys[pos] );
}

// finds the node in which the item should be stored
template <class T>
Node<T>* Node<T>::find_node( T item )
{
  if ( is_leaf() ) return this;
  int pos = find_pos( item );
  if ( equals_item( pos, item ) )
    return this;
```

```
  else
    return children[pos]->find_node( item );
}

template <class VEC>
VEC subseq( VEC vec, int start, int end )
{
  int size = ( vec.size() == 0 ) ? 0 : end - start;
  VEC result( size );
  for ( int i = 0; i<size; i++ )
    result[i] = vec[i + start];
  return result;
}

// if necessary, split the node.  Returns 0 or a new root
template <class T>
Node<T>* balance( Node<T>* node )
{
  int n = node->keys.size();
```

```
if ( n <= b ) return 0;
T median = node->keys[a];
// create a new node
Node<T>* node2 = new Node<T>( node->parent );
node2->keys = subseq( node->keys, a+1,
                      node->keys.size() );
node2->children = subseq( node->children, a+1,
                          node->children.size() );
for ( int i=0; i<node2->children.size(); i++ )
  node2->children[i]->parent = node2;
// handle node
node->keys = subseq( node->keys, 0, a );
node->children = subseq( node->children, 0, a+1 );

Node<T>* parent = node->parent;
if ( parent == 0 )  // split the root!
{
  Node<T>* root = new Node<T>( 0 );
  root->keys.push_back( median );
```

```
    root->children.push_back( node );
    root->children.push_back( node2 );
    node->parent = root;
    node2->parent = root;
    return root;
  }
  // otherwise: insert in parent
  int pos = 0;
  while ( parent->children[pos] != node ) pos++;
  parent->keys.insert( parent->keys.begin() + pos, median );
  parent->children.insert( parent->children.begin()+pos+1, node2 );
  // recursive call;
  return balance( parent );
}

template <class T>
void show( Node<T> *node )
{
  cout << node << ": ";
```

```
  if ( node->children.size() > 0 )
  {
    cout << node->children[0];
    for ( int i=0; i<node->keys.size(); i++ )
      cout << " |" << node->keys[i] << "| "
           << node->children[i + 1];
  }
  else
    for ( int i=0; i<node->keys.size(); i++ )
      cout << node->keys[i] << " ";
  cout << endl;
  for ( int i=0; i<node->children.size(); i++ )
    show( node->children[i] );
}

// we could work with a root pointer, but for later use it is
// better to wrap it into a class

template <class T>
```

```
class abTree
{
public:
  abTree() { root = new Node<T>( 0 ); }
  void insert( T item );
  bool find( T item );
  void show() { ::show( root ); }
private:
  Node<T>* root;
};

template <class T>
void abTree<T>::insert( T item )
{
  Node<T>* node = root->find_node( item );
  int i = node->find_pos( item );
  if ( node->equals_item( i, item ) )
    node->keys[i] = item;
  else
```

```
    {
      node->keys.insert( node->keys.begin()+i, item );
      Node<T>* new_root = balance( node );
      if ( new_root ) root = new_root;
    }
}

template <class T>
bool abTree<T>::find( T item )
{
  Node<T>* node = root->find_node( item );
  int i = node->find_pos( item );
  return node->equals_item( i, item );
}

int main()
{
  abTree<int> tree;
  // insertion demo
```

```
for ( int i=0; i<5; i++ )
{
  tree.insert( i );
  tree.show();
}
// testing insertion and retrieval
int n = 10;
for ( int i=0; i<n; i++ )
  tree.insert( i * i );
cout << endl;
tree.show();
for ( int i=0; i<2*n; i++ )
  cout << i << " " << tree.find( i ) << endl;
// performance test
//abTree<int> set;
set<int> set;  // should be faster
int nn = 1000000;
for ( int i=0; i<nn; i++ )
  set.insert( i * i );
```

```
  for ( int i=0; i<nn; i++ )
    set.find( i * i );
}
```

Bemerkung:

- Die Datensätze sind in allen Knoten gespeichert (Variante 2).
- Die Einfüge-Operation spaltet Knoten auf, wenn sie zu groß werden (mehr als b Schlüssel).
- Die Lösch-Operation ist nicht implementiert. Hier müssen Knoten vereinigt werden, wenn sie weniger als a Schlüssel enthalten.
- Die Effizienz ist zwar nicht schlecht, kommt aber nicht an die Effizienz der `set`-Implementation der STL heran. Ein wesentlicher Grund dafür ist die Verwendung variabel langer Vektoren zum Speichern von Schlüsseln und Kindern.

Literatur

Für eine weitergehende Darstellung von ausgeglichenen Bäumen sei auf das Buch „Grundlegende Algorithmen“ von Heun verwiesen. Noch mehr Informationen findet man im Buch „Introduction to Algorithms“ von Cormen, Leiserson, Rivest und Stein.

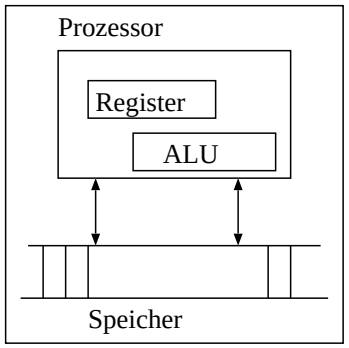
Bzw. besuchen Sie die Vorlesung ,,Algorithmen und Datenstrukturen”.

Beispiel: Logiksimulator

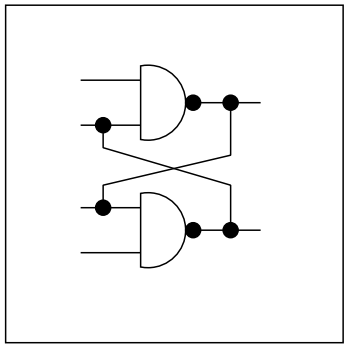
Zum Abschluss wollen wir ein größeres, zusammenhängendes Beispiel behandeln: Die Simulation von digitalen Schaltungen.

Simulation komplexer Systeme

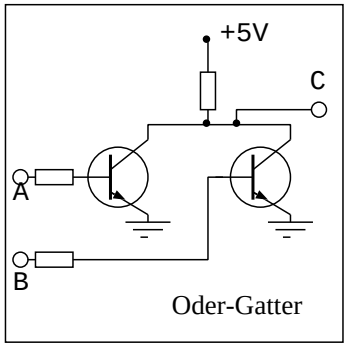
Ein Computer ist ein komplexes System, das auf verschiedenen Skalen modelliert werden kann.



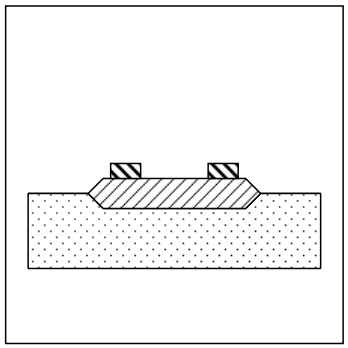
Aufbau aus Baugruppen: Prozessor, Speicher, Busse, Caches, Aufbau der Prozessoren aus Registern, ALU, usw.



Aufbau aus logischen Grundelementen: Und-Gatter, Oder-Gatter, Nand-Gatter, . . .
Zu jeder Zeit gibt es zwei mögliche Zustände auf jeder Leitung: $high$ und low. Modellierung durch boolsche Ausdrücke, zeitdiskrete, ereignisgesteuerte Simulation.



Die Gatter können wiederum aus Transistoren aufgebaut werden.
Spannungsverlauf an einem Punkt kontinuierlich. Modellierung durch Systeme gewöhnlicher Differentialgleichungen.



Aufbau eines Chips durch Schichten unterschiedlich dotierter Halbleiter.
Modellierung der Ladungsdichtenverteilung durch partielle Differentialgleichungen.

Wir beschäftigen uns hier mit der Modellierung und Simulation auf der Gatterebene.

Grundbausteine digitaler Schaltungen

Schaltungen auf der Logikebene sind aus digitalen Grundelementen aufgebaut.

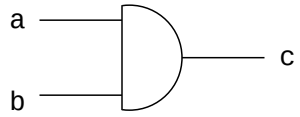
Die logischen Grundelemente besitzen einen oder mehrere Eingänge, sowie einen Ausgang.

Digital bedeutet, dass an einem Ein- bzw. Ausgang nur zwei verschiedene Spannungswerte vorkommen können: $high$ bzw. 1 oder low bzw. 0.

In Abhängigkeit der Belegung der Eingänge liefert der Baustein einen bestimmten Ausgangswert.

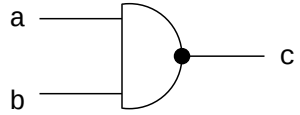
Hier sind die Grundschaltungen mit ihren Wahrheitstabellen:

And



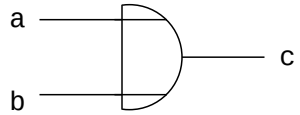
a	b	c
1	1	1
1	0	0
0	1	0
0	0	0

Nand



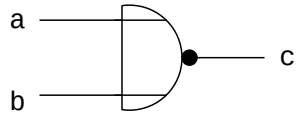
a	b	c
1	1	0
1	0	1
0	1	1
0	0	1

Or



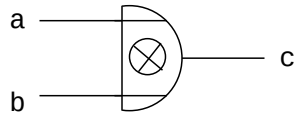
a	b	c
1	1	1
1	0	1
0	1	1
0	0	0

Nor



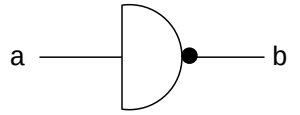
a	b	c
1	1	0
1	0	0
0	1	0
0	0	1

Exor



a	b	c
1	1	0
1	0	1
0	1	1
0	0	0

Inverter



a	b
1	0
0	1

Bemerkung: Durch Kombination mehrerer Nand-Gatter können alle anderen Gatter realisieriert werden.

Reale Gatter

Die digitalen Grundbausteine werden durch elektronische Schaltungen realisiert, die vollständig auf einem Halbleiterchip aufgebaut sind.

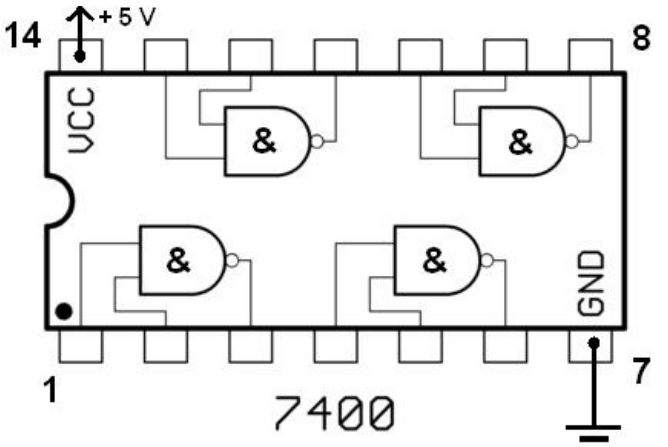
Im Prinzip besteht ein Mikroprozessor aus Millionen solcher Gatter auf einem einzigen Chip. Da sehr viele Gatter auf einem Chip sind spricht man von *very large scale integration* (VLSI).

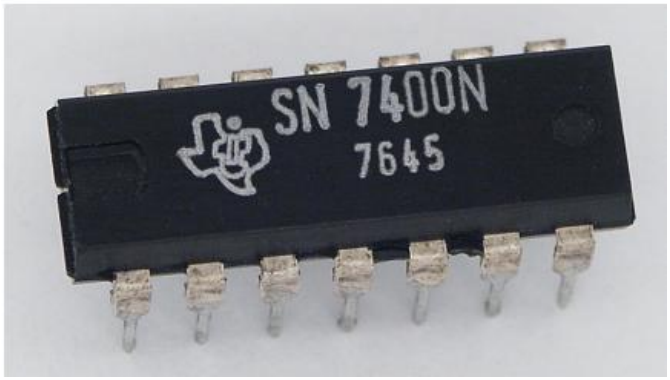
Es gibt aber auch einzelne (oder wenige) Gatter auf einem Chip, man spricht von *small scale integration* (SSI).

Bei den realen Gattern unterscheidet man verschiedene *Schaltungsfamilien*. Bausteine gleicher Familien können untereinander beliebig verschaltet werden, bei Bausteinen unterschiedlicher Familien ist dies nicht unbedingt möglich.

Eine der bekanntesten Familien digitaler Schaltglieder ist TTL (*transistor-transistor logic*). In einem Plastikgehäuse ist der Chip mit den Anschlüssen untergebracht.

So enthält der Baustein mit der Nummer 7400 vier Nand-Gatter:

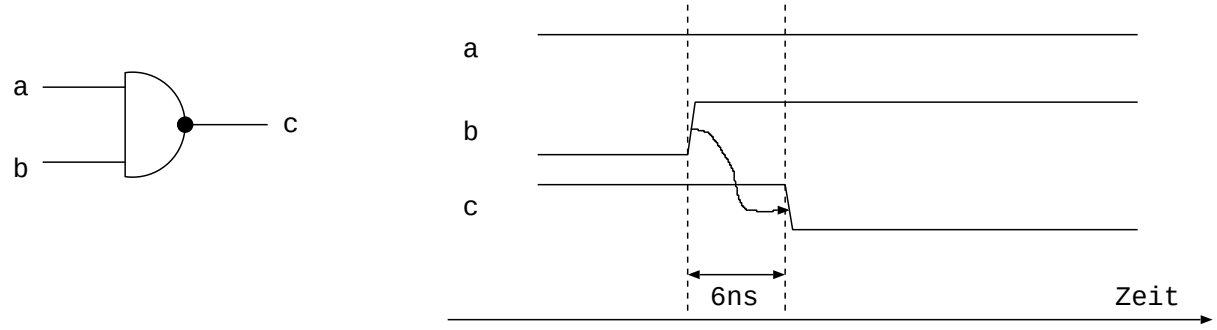




Bei TTL entspricht der logische Wert $high$ einer Spannung von $2 \ldots 5$ Volt, der Wert low $0 \ldots 0.4$ Volt.

Die Bausteine haben ausserdem ein *dynamisches Verhalten*, d. h. eine Änderung der

Eingangsspannung macht sich erst mit einer gewissen Verzögerung am Ausgang bemerkbar:



Schaltnetze

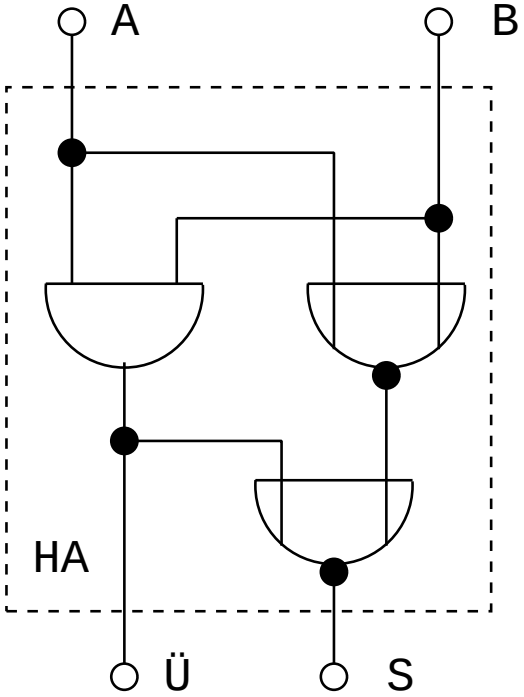
Ein *digitales Schaltnetz* ist eine Verknüpfung von logischen Grundelementen mit n Eingängen und m Ausgängen, wobei die Belegung der Ausgänge *nur* eine Funktion der Eingänge ist (abgesehen von der Verzögerung).

Ein Schaltnetz hat kein Gedächtnis, die Belegung der Ausgänge hängt nicht von früheren Belegungen der Eingänge ab.

Als Beispiel betrachten wir die Addition von Binärzahlen:

A3-A0	0	1	1	1	0
B3-B0	0	0	1	1	1
Übertrag	1	1	1		
Ergebnis	1	0	1	0	1

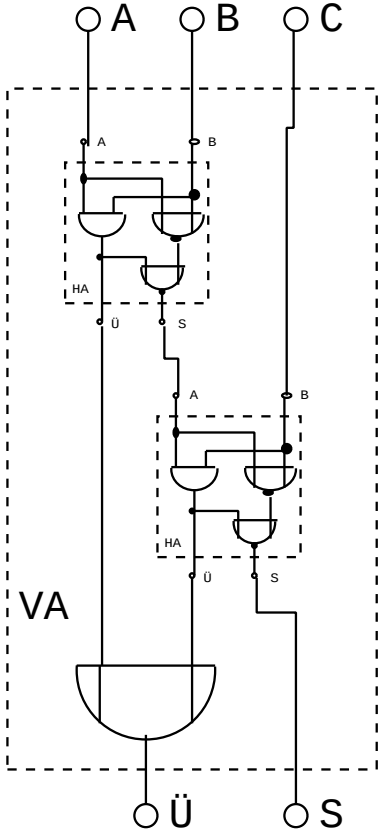
Halbaddierer: Addiert eine Binärstelle ohne Übertrag.



Wahrheits-
tabelle

A	B	S	Ü
1	1	0	1
1	0	1	0
0	1	1	0
0	0	0	0

Volladdierer: Addiert eine Binärstelle mit Übertrag.

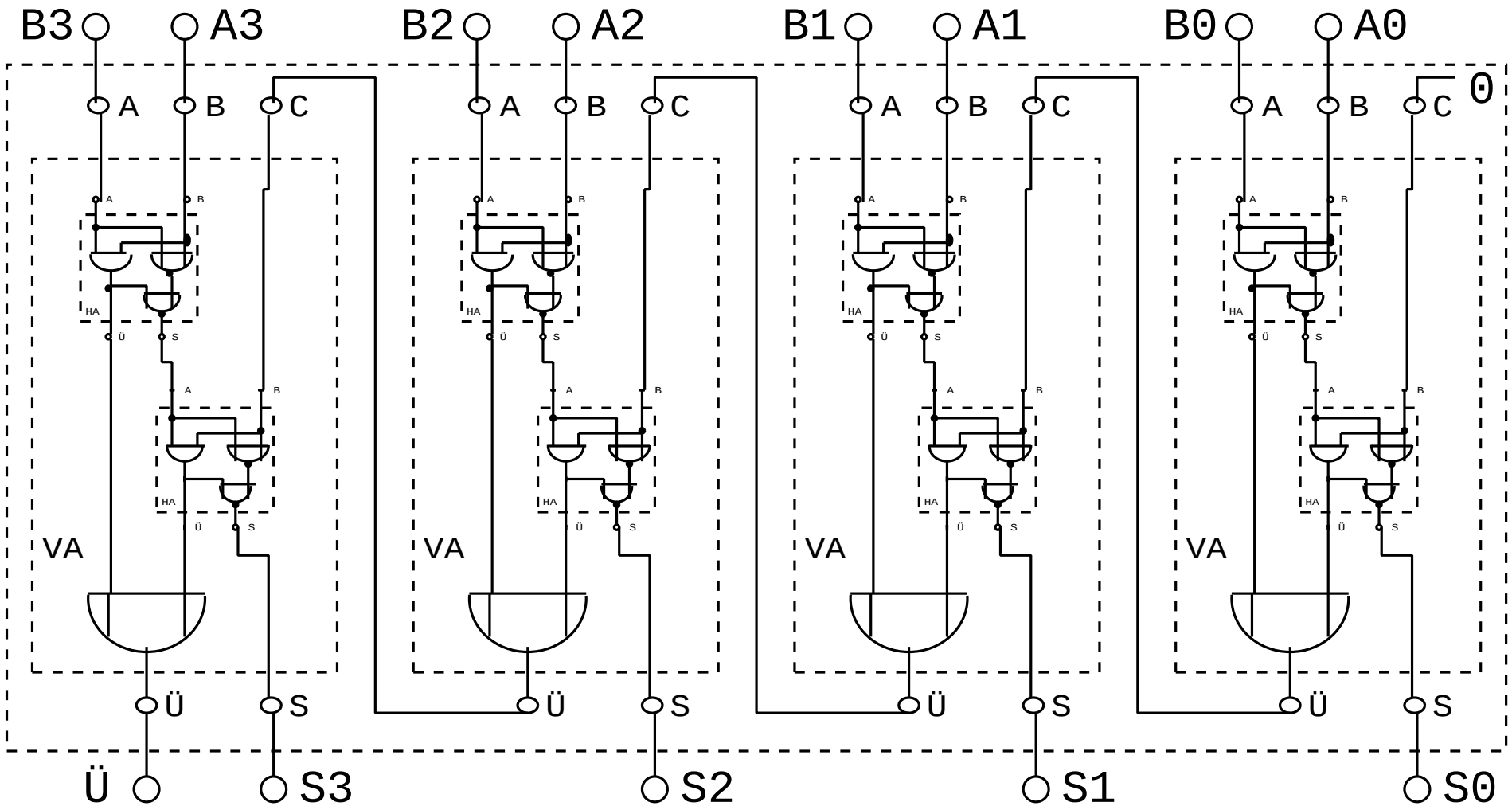


C	A	B	S	Ü
1	1	1	1	1
1	1	0	0	1
1	0	1	0	1
1	0	0	1	0
0	1	1	0	1
0	1	0	1	0
0	0	1	1	0
0	0	0	0	0

gerade Zahl von Einsen?

mehr als eine Eins?

4-Bit-Addierer mit durchlaufendem Übertrag

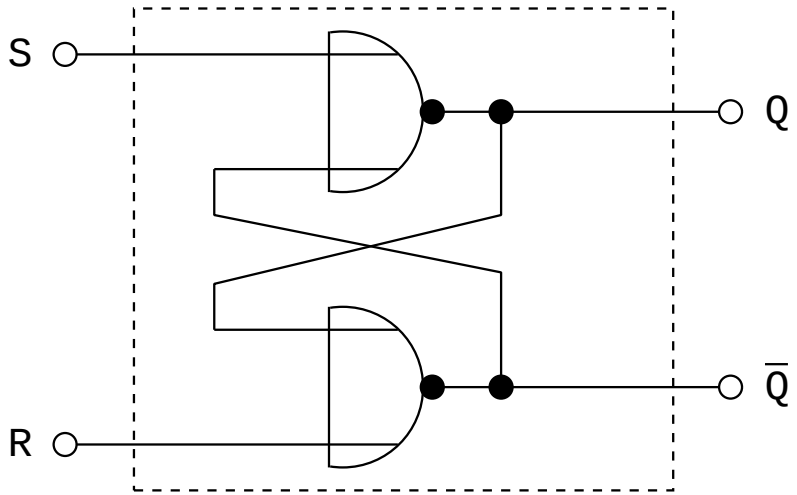


Schaltwerke

Ein *digitales Schaltwerk* ist eine Verknüpfung von Grundbausteinen mit n Eingängen und m Ausgängen, deren Belegung der Ausgänge von den Eingängen und internen Zuständen abhängt.

Die internen Zustände ergeben sich aus früheren Belegungen der Eingänge.

Wir betrachten als simpelstes Beispiel ein *SR-Flipflop* aus zwei Nor-Gattern:



Nor-Gatter

a	b	c
1	1	0
1	0	0
0	1	0
0	0	1

SR-Flipflop

S	R	Zust	Q	$\overline{Q}$	Zust
1	1	-	0	0	G
0	1	G	1	0	I
0	0	I	1	0	I
1	0	I	0	1	II
1	0	G	0	1	II
0	0	II	0	1	II
0	1	II	1	0	I

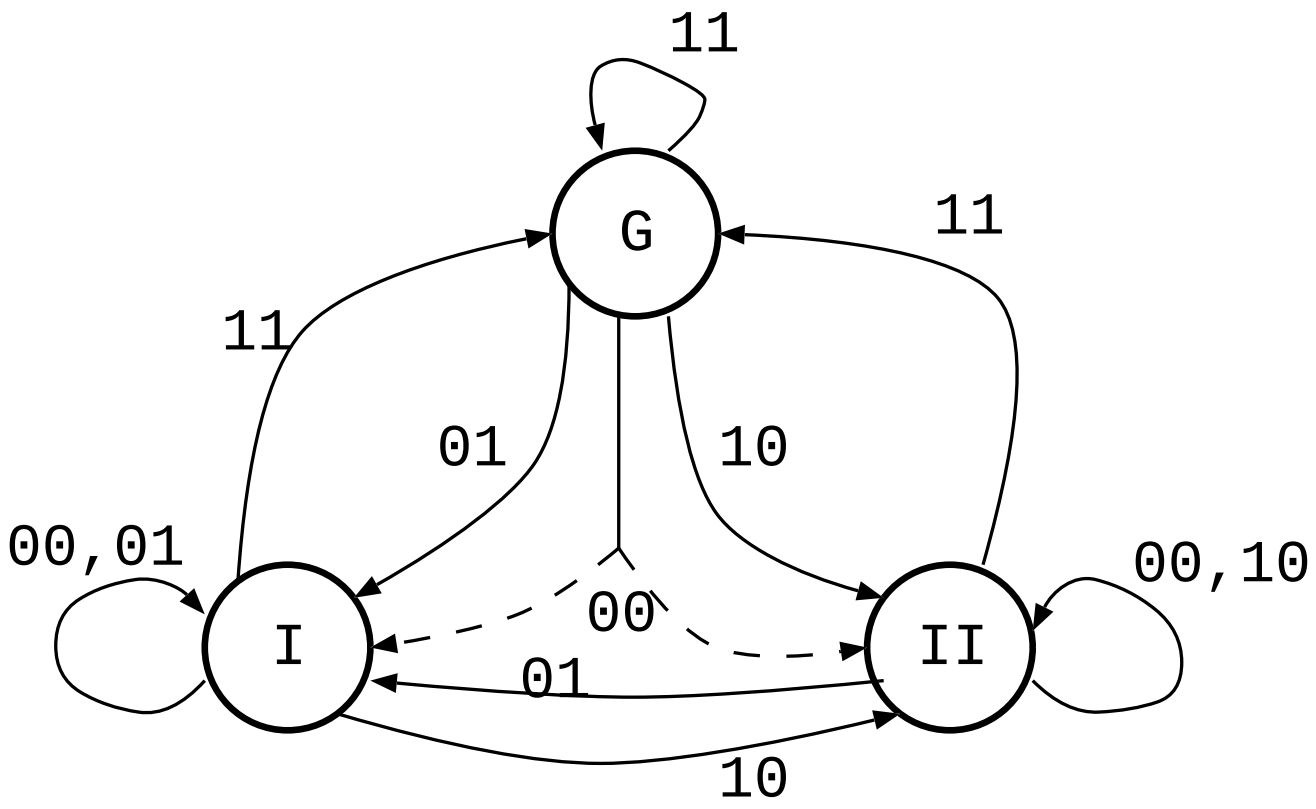
Die Eingangskombination $S = 0$, $R = 0$ hat verschiedene Ausgangsbelegungen zur Folge, je nach dem welche Signale vorher anlagen!

Bei zwei Ausgängen gibt es vier mögliche Kombinationen.

Die Kombination $Q = 1$, $\bar{Q} = 1$ ist nicht möglich, da eine 1 am Ausgang eines Nor-Gatters zweimal 0 am Eingang braucht, und das ist ein Widerspruch.

Die verbleibenden drei Zustände sind G ($Q = 0$, $\bar{Q} = 0$), I ($Q = 1$, $\bar{Q} = 0$) und II ($Q = 0$, $\bar{Q} = 1$) aus der Tabelle oben.

Aus der Tabelle oben erhalten wir das folgende Zustandsübergangsdiagramm:



Weitere Varianten von Flipflops können dazu benutzt werden um Speicher oder Zähler zu bauen.

Der Simulator

Die Zeit, die eine bestimmte Operation im Rechner benötigt, bestimmt sich aus den Gatterlaufzeiten der Schaltung die die Operation realisiert (siehe z. B. Addierer).

Der Schaltungsdesigner muss überprüfen ob die Schaltung in der Lage ist, die geforderte Operation innerhalb einer bestimmten Zeit (Takt) zu berechnen.

Dazu benutzt er einen Logiksimulator, der die Schaltung im Rechner nachbildet und simuliert. Er kann damit die Logikpegel an jedem Punkt der Schaltung über die Zeit verfolgen.

Die offensichtlichen Objekte in unserem System sind:

- Logische Grundbausteine mit ihrem Ein-/Ausgabeverhalten und Verzögerungszeit.
- Drähte, die Aus- und Eingänge der Gatter miteinander verbinden.

Eine andere Sache ist nicht sofort offensichtlich:

- In der Realität arbeiten alle Gatter simultan und unabhängig voneinander.
- In einem C++ Programm wird zu einer Zeit die Methode genau eines Objektes (=Gatter) ausgeführt. In welcher Reihenfolge sollen dann die Objekte bearbeitet werden?
- Wenn sich am Eingang eines Gatters nichts ändert, so ändert sich auch am Ausgang nichts. Um Arbeit zu sparen sollte also nur dort gearbeitet werden „wo sich etwas tut“.

Die Problematik löst man dadurch, dass man *Ereignisse*, wie etwa

„Eingang 1 von Gatter x geht in den Zustand $high$“

einführt. Das Eintreten eines Ereignisses löst dann wieder weitere Ereignisse bei anderen Objekten (Drähten oder Gattern) aus.

Die Koordination der Ereignisse wird von einem Objekt der Klasse `Simulator` übernommen.

Diese Art der Simulation bezeichnet man auch als *ereignisgesteuerte Simulation* (*discrete event simulation*).

Die Simulation als Ganzes gliedert sich in zwei Abschnitte:

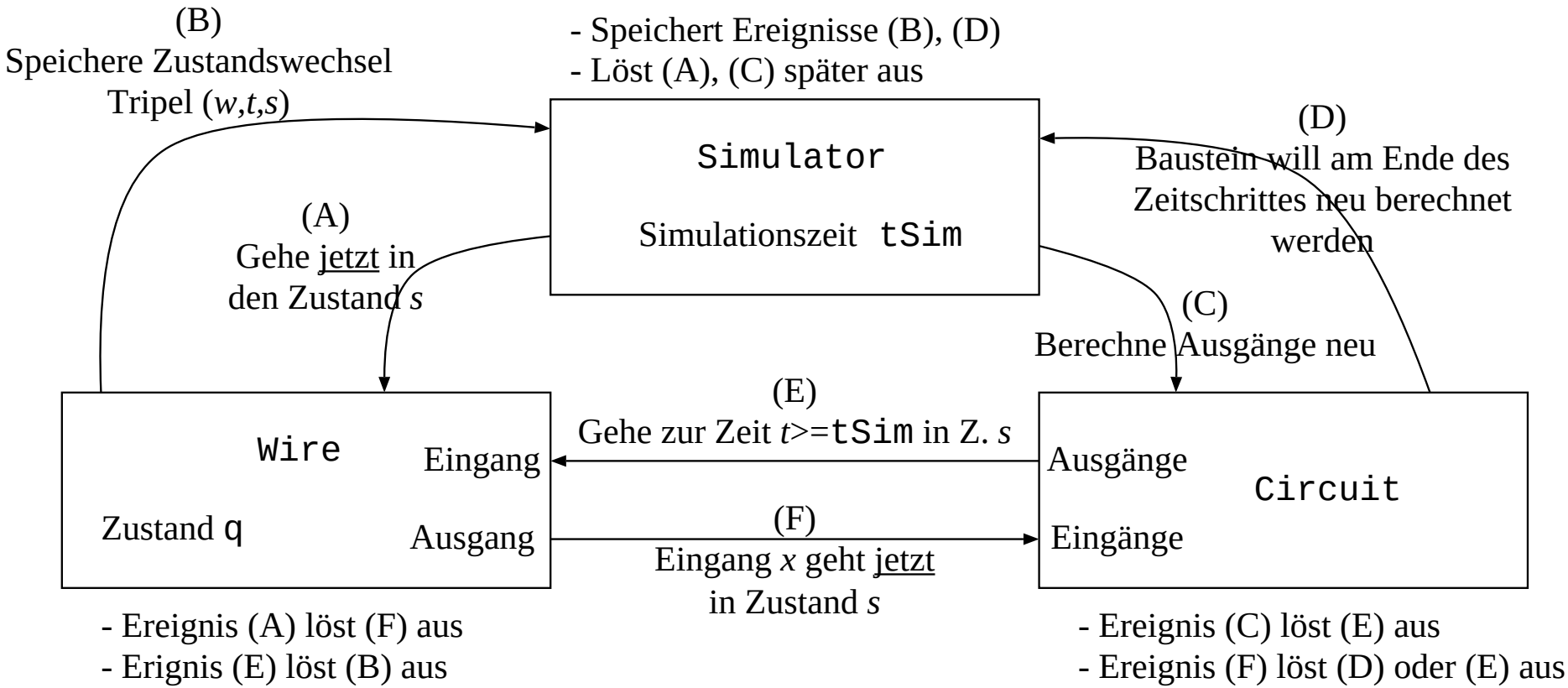
- Eine Aufbauphase, in der wir die zu simulierende Logikschaltung definieren:
  - Es werden Bausteine erzeugt. Alle Bausteine werden von der Schnittstellenklasse `Circuit` abgeleitet.
  - Die Verbindungsdrähte sind vom Typ `Wire` (eine konkrete Klasse). Drähte sind Punkt-zu-Punkt Verbindungen. Der Eingang eines Drahtes wird an den Ausgang eines Bausteins und der Ausgang eines Drahtes an den Eingang eines Bausteins angeschloßen. Verzweigungen werden durch einen speziellen Baustein (mit einem Eingang und zwei Ausgängen) realisiert.

- Die eigentliche Simulation. In der Aufbauphase werden erste Ereignisse generiert, die dann die weiteren Ereignisse auslösen.

  Die Simulationsphase geht in diskreten *Zeitschritten* vor und wird von einem Objekt der Klasse `Simulator` koordiniert.

  Die Klasse `Simulator` definiert eine globale Zeit $t_{Sim} \in \mathbb{N}$. Die Simulation beginnt zum Zeitpunkt $t_{Sim} = 0$.

Betrachten wir die Ereignisse, die zwischen den Objekten ausgetauscht werden:



Der Algorithmus für den Simulator lautet damit:

$t_{Sim} = 0$

Solange $t_{Sim} \leq t_{End}$

- $\forall$ gespeicherten Zustandswechsel (w, t, s) mit $t = t_{Sim}$: löse (A) aus.
- $\forall$ Bausteine deren Eingang sich im Zeitschritt t_{Sim} geändert hat: Berechne Baustein neu.
- $t_{Sim} = t_{Sim} + 1$

Draht

Die Klassendefinition für den Draht:

```
// moegliche Zustaende
enum State { low, high, unknown };

class Wire
{
public:
  // Draht im Zustand unknown erzeugen
  Wire();

  // aktuellen Zustand auslesen
  State GetState();

  // (E): Zur Zeit t soll Zustand s werden
```

```
  void ChangeState( int t, State s );

  // (A): wechsle jetzt in neuen Zustand
  void Action( State s );

  // Eingang des Drahtes an Ausgang i des Bausteins c
  // anschliessen
  void ConnectInput( Circuit& cir, int i );

  // Ausgang des Drahtes an Eingang i des Bausteins c
  // anschliessen
  void ConnectOutput( Circuit& cir, int i );

private:
  State q;        // der Zustand
  Circuit* c;     // Baustein am Ausgang des Drahtes
  int pin;        // pin des Bausteins
```

```
};
```

Der Zustand kann auch unbekannt sein, um feststellen zu können, ob die Schaltung korrekt initialisiert wird.

Der Konstruktor erzeugt den Draht im Zustand `unknown`.

Die nächsten drei Methoden werden in der Simulationsphase verwendet.

Die letzten beiden Methoden werden in der Aufbauphase verwendet. Der Draht merkt sich an Baustein und Eingang des Bausteins an den er angeschlossen wurde.

. . . und die Implementierung der Methoden

```
Wire::Wire()
{
  // Initialisiere mit unbekanntem Zustand
  q = unknown;
}
```

```
inline State Wire::GetState()
{
  return q;
}

void Wire::ChangeState( int t, State s )
{
  Sim.StoreWireEvent( *this, t, s );
}

void Wire::Action( State s )
{
  if ( s == q ) return;      // nix zu tun
  q = s;                     // neuer Zustand
  // Nachricht an angeschlossen Baustein
  c->ChangeInput( q, pin );
}

void Wire::ConnectInput( Circuit& cir, int i )
```

```
{
  // Merke NICHT an wen ich angeschlossen bin
  // aber Baustein muss mich kennen.
  cir.ConnectOutput( *this, i );
}

void Wire::ConnectOutput( Circuit& cir, int i )
{
  // Merke Baustein, an den der Ausgang angeschlossen ist
  c = &cir;
  pin = i;
  // Rueckverbindung Baustein an Draht
  c->ConnectInput( *this, pin );
}
```

Das Simulatorobjekt `Sim` ist ein globales Objekt, das alle kennen.

Es gibt nur ein Simulatorobjekt. Wenn es keinen Sinn macht mehr als ein Objekt von einer Klasse zu instanzieren, nennt man dies ein *Singleton*.

Bausteine

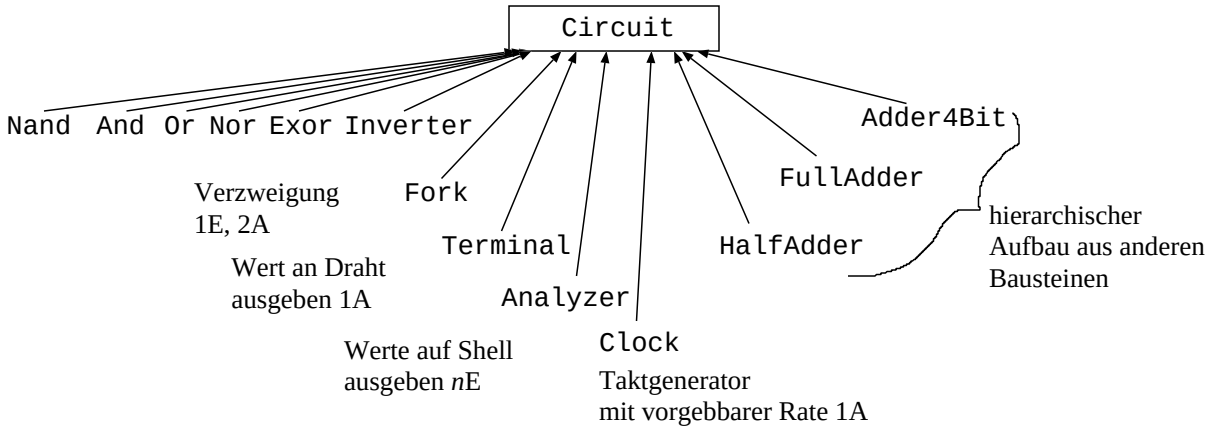
Die Schnittstellenbasisklasse für Bausteine:

```
class Circuit
{
public:
  // virtual destructor
  virtual ~Circuit();

  // (F): Eingang wechselt Zustand
  virtual void ChangeInput( State s, int pin ) = 0;

  // (C): Ausgang neu berechnen
  virtual void Action() = 0;

  // verdrahte Eingang
```

```
    virtual void ConnectInput( Wire& w, int pin ) = 0;

    // verdrahte Ausgang
    virtual void ConnectOutput( Wire& w, int pin ) = 0;
};

Circuit::~Circuit() {}
```

Die `Connect...`-Funktionen werden von der entsprechenden Funktion der Klasse `Wire` aufgerufen.

. . . und die zur Zeit existierenden abgeleiteten Klassen



Hier die abgeleitete Klasse für das Nand-Gatter:

```
class Nand : public Circuit
{
public:
  // Konstruktor
  Nand();

  // default destructor ist OK
  ~Nand();
```

```
  // Eingang wechselt zur aktuellen Zeit den Zustand
  virtual void ChangeInput( State s, int pin );

  // berechne Gatter neu und benachrichtige Draht
  // am Ausgang
  virtual void Action();

  // verdrahte Eingang
  virtual void ConnectInput( Wire& w, int pin );

  // verdrahte Ausgang
  virtual void ConnectOutput( Wire& w, int pin );

private:
  Wire* a;              // Eingang 1
  Wire* b;              // Eingang 2
```

```
    Wire* c;              // Ausgang
    bool actionFlag;      // merke ob bereits aktiviert
};
```

Das Gatter merkt sich, welche Drähte an Ein- und Ausgängen angeschlossen sind.

und die Methoden des Nand-Gatters:

```
Nand::Nand()
{
    a = b = c = 0;  // nix angschlossen
    actionFlag = false;
}

Nand::~Nand() {}

void Nand::ChangeInput( State s, int pin )
{
```

```
    // Sorge dafuer, dass Gatter neu berechnet wird
    if ( !actionFlag )
    {
      Sim.StoreCircuitEvent( *this );
      actionFlag = true;
    }
}

void Nand::Action()
{
    // Lese Eingangssignale
    State A = a->GetState();
    State B = b->GetState();
    State Output = unknown;

    // Wertetabelle
    if ( A == high && B == high ) Output = low;
```

```
  if ( A == low || B == low ) Output = high;

  // Setze Draht
  if ( c != 0 )
    c->ChangeState( Sim.GetTime() + 3, Output );

  // erlaube neue Auswertung
  actionFlag = false;
}

void Nand::ConnectInput( Wire& w, int pin )
{
  // Wird von Connect-Funktion des Drahtes aufgerufen
  if ( pin == 0 ) a = &w;
  if ( pin == 1 ) b = &w;
}
```

```
void Nand::ConnectOutput( Wire& w, int pin )
{
  // Wird von Connect-Funktion des Drahtes aufgerufen
  c = &w;
}
```

Ein zweites Beispiel ist die Verzweigung `Fork.hh`:

```
class Fork : public Circuit
{
public:
  // Konstruktor
  Fork();

  // default destructor ist OK
  ~Fork();
```

```
  // Eingang wechselt zur aktuellen Zeit den Zustand
  virtual void ChangeInput( State s, int pin );

  // berechne Gatter neu und benachrichtige Draht
  // am Ausgang
  virtual void Action();

  // verdrahte Eingang
  virtual void ConnectInput( Wire& w, int pin );

  // verdrahte Ausgang
  virtual void ConnectOutput( Wire& w, int pin );

private:
  Wire* a;              // Eingang
  Wire* b;              // Ausgang 1
  Wire* c;              // Ausgang 2
```

```
};
```

Hier werden sofort Ereignisse bei den Drähten am Ausgang ausgelöst:

```
Fork::Fork()
{
  a = b = c = 0; // nix angschlossen
}

Fork::~Fork() {}

void Fork::ChangeInput( State s, int pin )
{
  // Leite Eingang SOFORT an beide Ausgaenge weiter
  if ( b != 0 ) b->ChangeState( Sim.GetTime(), s );
  if ( c != 0 ) c->ChangeState( Sim.GetTime(), s );
}
```

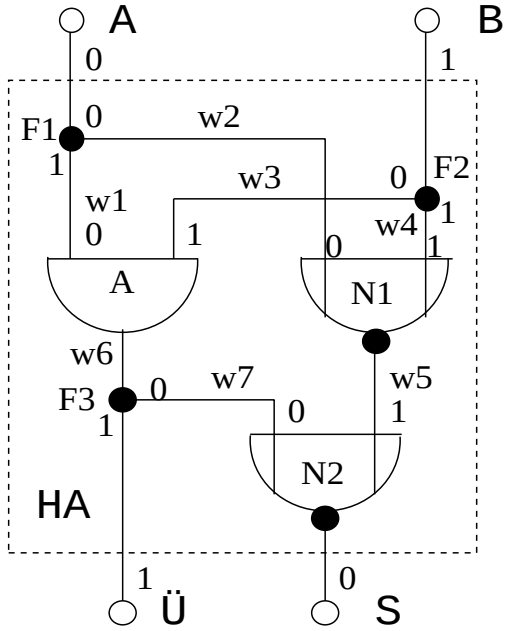
```
void Fork::Action()
{
  // nix zu tun
}

void Fork::ConnectInput( Wire& w, int pin )
{
  // Wird von Connect-Funktion des Drahtes aufgerufen
  a = &w;
}

void Fork::ConnectOutput( Wire& w, int pin )
{
  // Wird von Connect-Funktion des Drahtes aufgerufen
  if ( pin == 0 ) b = &w;
  if ( pin == 1 ) c = &w;
```

```
}
```

`HalfAdder` ist ein Beispiel für einen zusammengesetzten Baustein:

```
class HalfAdder : public Circuit
{
public:
  // Konstruktor
  HalfAdder();

  // destruktor
  ~HalfAdder();

  // Eingang wechselt zur aktuellen Zeit den Zustand
  virtual void ChangeInput( State s, int pin );

  // berechne Gatter neu und benachrichtige Draht
```

```
  // am Ausgang
  virtual void Action();

  // verdrahte Eingang
  virtual void ConnectInput( Wire& w, int pin );

  // verdrahte Ausgang
  virtual void ConnectOutput( Wire& w, int pin );

private:
  Wire w1, w2, w3, w4, w5, w6, w7; // lokale Draehte
  And A;                           // Und Gatter
  Nor N1, N2;                      // sowie zwei Nor Gatter
  Fork F1, F2, F3;                 // und drei Verzweigungen
};
```



. . . und die Methoden:

```
HalfAdder::HalfAdder()
{
  w1.ConnectInput( F1, 1 );
  w1.ConnectOutput( A, 0 );
  w2.ConnectInput( F1, 0 );
```

```
  w2.ConnectOutput( N1, 0 );
  w3.ConnectInput( F2, 0 );
  w3.ConnectOutput( A, 1 );
  w4.ConnectInput( F2, 1 );
  w4.ConnectOutput( N1, 1 );
  w5.ConnectInput( N1, 0 );
  w5.ConnectOutput( N2, 1 );
  w6.ConnectInput( A, 0 );
  w6.ConnectOutput( F3, 0 );
  w7.ConnectInput( F3, 0 );
  w7.ConnectOutput( N2, 0 );
}

HalfAdder::~HalfAdder() {}

void HalfAdder::ChangeInput( State s, int pin )
{
```

```
  if ( pin == 0 ) F1.ChangeInput( s, 0 );
  if ( pin == 1 ) F2.ChangeInput( s, 0 );
}

void HalfAdder::Action() {}

void HalfAdder::ConnectInput( Wire& w, int pin )
{
  // Wird von Connect-Funktion des Drahtes aufgerufen
  if ( pin == 0 ) F1.ConnectInput( w, 0 );
  if ( pin == 1 ) F2.ConnectInput( w, 0 );
}

void HalfAdder::ConnectOutput( Wire& w, int pin )
{
  // Wird von Connect-Funktion des Drahtes aufgerufen
  if ( pin == 0 ) N2.ConnectOutput( w, 0 );
```

```
    if ( pin == 1 ) F3.ConnectOutput( w, 1 );
}
```

Simulator

Schließlich der Simulator

```
// Simulator, Singleton
class Simulator
{
public:
  // Konstruktor
  Simulator();

  // aktuelle Zeit auslesen
  int GetTime();

  // (B): Draht w wird zur Zeit t in Zustand s wechseln
  void StoreWireEvent( Wire& w, int t, State s );

  // (D): Baustein c soll zur aktuellen Zeit neu
  // berechnet werden
```

```
  void StoreCircuitEvent( Circuit& c );

  // Starte Simulation bei Zeit 0
  void Simulate( int end );

private:
  struct WireEvent
  { // Eine lokale Struktur
    WireEvent();  // fuer Ereignis "Zustandswechsel"
    WireEvent( Wire& W, int T, State S );
    Wire* w;
    int t;
    State s;
    bool operator<( WireEvent we );
    void print( std::ostream& stm ) {
      stm << "(WE: " << t << " " << w << " " << s << std::endl;
    }
  };
```

```
  int time;
  MinPriorityQueue<WireEvent> pq; // Fuer (B)-Ereignisse
  Queue<Circuit*> q;              // Fuer (D)-Ereignisse
};

// Globale Variable vom Typ Simulator (Singleton).
// Wird von allen Bausteinen und Draehten benutzt!
Simulator Sim;
```

. . . und seine Methoden:

```
// Methoden fuer die geschachtelte Klasse
Simulator::WireEvent::WireEvent() {
  w = 0; t = 0; s = unknown;
}

Simulator::WireEvent::WireEvent( Wire& W, int T, State S ) {
  w = &W; t = T; s = S;
}
```

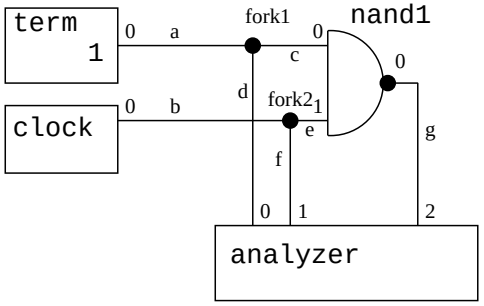
```
bool Simulator::WireEvent::operator<( WireEvent we )
{
  if ( t < we.t ) return true;
  if ( t == we.t &&
       ( reinterpret_cast<unsigned long int>( w ) <
         reinterpret_cast<unsigned long int>( we.w ) ) )
    return true;
  return false;
}

// Konstruktor
Simulator::Simulator() { time = 0; }

int Simulator::GetTime() { return time; }

void Simulator::StoreWireEvent( Wire& w, int t, State s )
{
  pq.push( WireEvent( w, t, s ) );
}
```

```
void Simulator::StoreCircuitEvent( Circuit& c )
{
  q.push( &c );
}

void Simulator::Simulate( int end )
{
  WireEvent we;

  while ( time <= end )
  {
    // Alle Draehte fuer die aktuelle Zeit
    while ( !pq.empty() )
    {
      we = pq.top();              // kleinster Eintrag
      // alle Zustaende fuer Zeitschritt OK
      if ( we.t > time ) break;
      pq.pop();                   // entferne Eintrag
```

```
      (we.w)->Action( we.s );    // neuer Zustand
    }

    // Berechne Bausteine zur aktuellen Zeit neu
    while ( !q.empty() )
    {
      (q.front())->Action();
      q.pop();
    }

    // Zeitschritt fertig
    time = time + 1;
  }
}
```

Beispiel: Nand-Gatter an Taktgenerator

Als Beispiel betrachten wir folgende Schaltung:



Dafür produziert der Simulator folgende Ausgabe:

```
0   U   U   U
1   U   1   1
2   U   1   1
3   U   1   1
4   0   1   1
5   0   1   1
6   0   1   1
7   0   1   1
8   0   1   1
```

9	0	1	1
10	0	1	1
11	0	0	1
12	0	0	1
13	0	0	1
14	1	0	1
15	1	0	1
16	1	0	1
17	1	0	1
18	1	0	1
19	1	0	1
20	1	0	1
21	1	1	1
22	1	1	1
23	1	1	1
24	0	1	1
25	0	1	1

```
26   0   1   1
27   0   1   1
28   0   1   1
29   0   1   1
30   0   1   1
31   0   0   1
32   0   0   1
33   0   0   1
```

Hier das zugehörige Hauptprogramm:

```
#include <iostream>
#include <cassert>
#include "DLL.hh"
#include "MinPriorityQueue.hh"
#include "MinPriorityQueueImp.cc"
#include "Queue.hh"
class Simulator; // forward declaration
class Wire;      // forward declaration
```

```
class Circuit;    // forward declaration
#include "Wire.hh"
#include "Circuit.hh"
#include "Simulator.hh"
#include "SimulatorImp.cc"
#include "WireImp.cc"
#include "Nand.hh"
#include "NandImp.cc"
#include "Terminal.hh"
#include "TerminalImp.cc"
#include "Fork.hh"
#include "ForkImp.cc"
#include "Analyzer.hh"
#include "AnalyzerImp.cc"
#include "Clock.hh"
#include "ClockImp.cc"

int main()
{
```

```
  Nand nand1; Analyzer analyzer( 3 );
  Fork fork1, fork2; Clock clock( 10, high );
  Wire a, b, c, d, e, f, g; Terminal term( high );

  a.ConnectInput( term, 0 );   a.ConnectOutput( fork1, 0 );
  b.ConnectInput( clock, 0 );  b.ConnectOutput( fork2, 0 );
  c.ConnectInput( fork1, 0 );  c.ConnectOutput( nand1, 0 );
  d.ConnectInput( fork1, 1 );  d.ConnectOutput( analyzer, 0 );
  e.ConnectInput( fork2, 0 );  e.ConnectOutput( nand1, 1 );
  f.ConnectInput( fork2, 1 );  f.ConnectOutput( analyzer, 1 );
  g.ConnectInput( nand1, 0 );  g.ConnectOutput( analyzer, 2 );

  Sim.Simulate( 33 );
}
```

Beispiel: 4-Bit Addierer

Es geht komplizierter, die 4 Bit Addition

```
#include <iostream>
#include <cassert>
#include "DLL.hh"
#include "MinPriorityQueue.hh"
#include "MinPriorityQueueImp.cc"
#include "Queue.hh"
class Simulator; // forward declaration
class Wire;      // forward declaration
class Circuit;   // forward declaration
#include "Wire.hh"
#include "Circuit.hh"
#include "Simulator.hh"
#include "SimulatorImp.cc"
#include "WireImp.cc"
#include "Nand.hh"
```

```
#include "NandImp.cc"
#include "And.hh"
#include "AndImp.cc"
#include "Nor.hh"
#include "NorImp.cc"
#include "Or.hh"
#include "OrImp.cc"
#include "Exor.hh"
#include "ExorImp.cc"
#include "Inverter.hh"
#include "InverterImp.cc"
#include "Fork.hh"
#include "ForkImp.cc"
#include "Terminal.hh"
#include "TerminalImp.cc"
#include "Analyzer.hh"
#include "AnalyzerImp.cc"
#include "Clock.hh"
#include "ClockImp.cc"
```

```
#include "HalfAdder.hh"
#include "HalfAdderImp.cc"
#include "FullAdder.hh"
#include "FullAdderImp.cc"
#include "Adder4Bit.hh"
#include "Adder4BitImp.cc"

int main()
{
  Adder4Bit adder;
  Analyzer analyzer( 5 );
  Terminal a3( low ), a2( high ), a1( high ), a0( high );
  Terminal b3( high ), b2( low ), b1( high ), b0( low );
  Terminal c0( low );

  Wire wa0, wa1, wa2, wa3;
  Wire wb0, wb1, wb2, wb3;
  Wire ws0, ws1, ws2, ws3;
  Wire wc0, wc4;
```

```
wc0.ConnectInput( c0, 0 );
wc0.ConnectOutput( adder, 8 );

wa0.ConnectInput( a0, 0 );
wa1.ConnectInput( a1, 0 );
wa2.ConnectInput( a2, 0 );
wa3.ConnectInput( a3, 0 );
wa0.ConnectOutput( adder, 0 );
wa1.ConnectOutput( adder, 1 );
wa2.ConnectOutput( adder, 2 );
wa3.ConnectOutput( adder, 3 );

wb0.ConnectInput( b0, 0 );
wb1.ConnectInput( b1, 0 );
wb2.ConnectInput( b2, 0 );
wb3.ConnectInput( b3, 0 );
wb0.ConnectOutput( adder, 4 );
wb1.ConnectOutput( adder, 5 );
```

```
    wb2.ConnectOutput( adder, 6 );
    wb3.ConnectOutput( adder, 7 );

    ws0.ConnectInput( adder, 0 );
    ws1.ConnectInput( adder, 1 );
    ws2.ConnectInput( adder, 2 );
    ws3.ConnectInput( adder, 3 );
    ws0.ConnectOutput( analyzer, 0 );
    ws1.ConnectOutput( analyzer, 1 );
    ws2.ConnectOutput( analyzer, 2 );
    ws3.ConnectOutput( analyzer, 3 );

    wc4.ConnectInput( adder, 4 );
    wc4.ConnectOutput( analyzer, 4 );

    Sim.Simulate( 40 );
  }
```

. . . und die Ausgabe:

```
 0  U  U  U  U  U
 1  U  U  U  U  U
 2  U  U  U  U  U
 3  U  U  U  U  U
 4  U  U  U  U  U
 5  U  U  U  U  U
 6  U  U  U  U  U
 7  U  U  U  U  U
 8  U  U  U  U  U
 9  U  U  U  U  U
10  U  U  U  U  U
11  U  U  U  U  U
12  U  U  U  U  U
13  U  U  U  U  1
14  U  U  U  U  1
15  U  U  U  U  1
16  U  U  U  U  1
```

17	U	U	U	U	1
18	U	U	U	U	1
19	U	U	U	0	1
20	U	U	U	0	1
21	U	U	U	0	1
22	U	U	U	0	1
23	U	U	U	0	1
24	U	U	U	0	1
25	U	U	0	0	1
26	U	U	0	0	1
27	U	U	0	0	1
28	U	U	0	0	1
29	U	U	0	0	1
30	U	U	0	0	1
31	1	0	0	0	1
32	1	0	0	0	1
33	1	0	0	0	1

34	1	0	0	0	1
35	1	0	0	0	1
36	1	0	0	0	1
37	1	0	0	0	1
38	1	0	0	0	1
39	1	0	0	0	1
40	1	0	0	0	1

Verschiedenes

In diesem Abschnitt wollen wir noch einige Punkte erwähnen, die für das Programmieren wichtig sind.

Rechtliches

Lizenzen

Software ist normalerweise urheberrechtlich geschützt. Das geht so weit, dass es nicht erlaubt ist, Software ohne Erlaubnis des Besitzers in irgendeiner Weise zu nutzen. Diese Erlaubnis wird durch verschiedene Lizenzen erteilt, die dem Benutzer mehr oder weniger Einschränkungen auferlegen.

Wichtige Lizenzen

- Public Domain: der Code ist völlig frei, alles ist erlaubt. Vorsicht: nach deutschem Recht kann ein Programmierer seine Rechte in dieser Weise nicht aufgeben!

- BSD-Lizenz (nach dem Betriebssystem BSD Unix benannt): schließt Verantwortung des Urhebers für Fehler aus, erlaubt dem Benutzer alles außer Verändern der Lizenz selbst.

- GNU GPL (GNU General Public License): erlaubt dem Benutzer privat alles, bei Weitergabe an Dritte muss er aber auf deren Anfrage hin auch die Quelltexte

unter der GPL nachliefern. Dasselbe muss er auch für angebundene Bibliotheken tun, die daher „kompatible“ Lizenzen haben müssen. Der Linux-Kernel steht unter der GPL.

- MPL (Mozilla Public License), QPL (Qt Public License): ähnlich GPL, aber dem Erstautor (Netscape, Trolltech, . . .) werden besondere Rechte eingeräumt.

- Akademische Lizenzen: Gebrauch im akademischen Bereich erlaubt, sonst besondere Erlaubnis nötig.

- Kommerzielle Lizenzen: Oft erhebliche Einschränkungen, normalerweise ist kein Quellcode enthalten.

Wer hat die Rechte an Software?

Generell hat der Schöpfer eines Werks das Urheberrecht. Allerdings besteht oft ein Vertrag, der die Nutzungsrechte dem Arbeitgeber überträgt. Im akademischen Bereich ist das oft die entsprechende Hochschule.

Wenn der Arbeitgeber nicht an einer Nutzung der Software interessiert sein sollte (er muss aber gefragt werden!), so können die Rechte nach einer Wartezeit an den Angestellten rückübertragen werden.

Softwarepatente

Leider werden vor allem in den USA Patente auf softwaretechnische Ideen erteilt, die eigentlich nicht patentierbar sein sollten (z. B. Fortschrittsbalken, One-Click-Shopping), weil sie „offensichtlich“ sind (d. h. viele Programmierer hätten dieselbe Lösung für dasselbe Problem gefunden) oder aber erst nach einer wissenschaftlichen Veröffentlichung patentiert wurden (z. B. RSA-Kryptographie). Auch wenn solche Patente in Europa bisher noch nicht gültig sind, so kann es natürlich Schwierigkeiten bei in die USA exportierter Software geben.

Software-Technik (Software-Engineering)

Werkzeuge

- Editor (muss Sprachelemente kennen, kann in großen Programmsystemen Definitionen finden, Teile separat kompilieren und testen)
- Versionsverwaltung: Rückverfolgung von Änderungen, Vereinigung der Änderungen verschiedener Programmierer, Release-Management; Unix: CVS, Subversion, Git, . . .
- Debugger: Beobachtung des laufenden Programms; Unix: gdb
- Profiling: Messen des Speicher- und Rechenzeitverbrauchs; Programme: valgrind (kommerziell: Purify), gprof;
- Testumgebung z. B. CppUnit.
- Entwicklungsumgebung (IDE): alles integriert

Beobachtungen aus der Praxis

- Große Programmpakete sind die Arbeit vieler Programmierer.
- Die Produktivität von Programmierern weist gigantische Unterschiede auf. Zwischen Durchschnitts- und Spitzenprogrammierer kann man mit etwa einer Größenordnung rechnen.
- Kleine Gruppen von Programmierern sind relativ gesehen am produktivsten. Insbesondere Zweiergruppen können sich gut ergänzen (pair programming).
- Es ist besonders effektiv, wenn ein Architekt ein Programmprojekt leitet. Dieser Architekt sollte
  - ein guter Programmierer sein,
  - sich mit vorhandenen Bibliotheken auskennen,
  - sich auch im Anwendungsbereich auskennen und

– Autorität besitzen.

- Als sehr gut hat sich inkrementelle Entwicklung erwiesen: man fängt mit einem Prototyp an, der sukzessive verfeinert wird. Es wird darauf geachtet, dass möglichst immer eine lauffähige Version des Programms zur Verfügung steht.

- Ambitionierte Konzepte brauchen (zu viel) Zeit. Dies kann im Wettlauf mit anderen das Ende bedeuten.

- Ein Großteil der Kosten entsteht oft bei der Wartung der Software. Schlechte Arbeit am Anfang (z. B. verfehlter Entwurf) kann hier verhältnismäßig große Kosten verursachen.

Was kann schiefgehen?

- Manchmal ist die fachliche Qualifikation von Personen in Leitungsfunktionen (Architekt, Manager) schlecht, was zu teilweise schwerwiegenden Fehlentscheidungen führt.

- Insbesondere sind Manager (und Programmierer ebenso) viel zu optimistisch mit Abschätzungen der Schwierigkeit eines Programmprojekts. Ninety-ninety rule (Tom Cargill, Bell Labs):

  > The first 90% of the code accounts for the first 90% of the development time. The remaining 10% of the code accounts for the other 90% of the development time.

- Einsparungen und Druck von oben können dazu führen, dass gerade die besten Programmierer gehen.

- Die Einstellung neuer Programmierer, um eine Deadline einzuhalten, kann die Arbeit noch weiter verzögern, weil diese eingearbeitet werden müssen.

Literatur: Frederick P. Brooks: *The Mythical Man-Month* (Klassiker)

Schlagworte

- Wasserfallmodell: Planung (*Systems Engineering*), Definition (*Analysis*), Entwurf (*Design*), Implementierung (*Coding*), Testen (*Testing*), Einsatz und Wartung (*Maintenance*)
- Agile Software-Entwicklung: XP (Extreme Programming, Pair programming), Scrum (Tägliche kurze Treffen, Sprints)
- Rational Unified Process: macht Wasserfallmodell iterativ, benutzt UML (Universal Modeling Language)
- XML: Datenaustauschformat, menschen- und maschinenlesbar, einfache Syntax der Form `<name>...</name>`. Wird leider beliebig kompliziert durch Schemas (DTD, XML Schema)

- Entwurfsmuster (Design Patterns): Die Idee ist es, wiederkehrende Muster in Programmen jenseits der Sprachelemente zu identifizieren und zu nutzen. Entwurfsmuster wurden bekannt durch das Buch: Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides: *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison Wesley, 1995, ISBN 0-201-63361-2.

Bemerkung: Vorsicht vor Schlagworten, wenn sie als Allheilmittel verkauft werden. Viele Programmieraufgaben sind fundamental schwierig, und es wird auf absehbare Zeit dafür keine einfache Lösung geben geben.

Wie werde ich ein guter Programmierer?

Inhalt der Vorlesung

- Grundlagen des Programmierens in C/C++
- Kennenlernen verschiedener Programmiertechniken (funktional, imperativ, objektorientiert, generisch)
- Erkennen der Bedeutung effizienter Algorithmen

Bemerkung: Auch wenn C/C++ gerade für Anfänger nicht die einfachste Sprache ist, sind grundlegende Kenntnisse darin unbedingt notwendig. Das von der Syntax her ähnliche Java und C/C++ decken einen großen Teil des Software-Markts ab.

Nicht behandelt

Die Vorlesung hat aber viele interessante Themen nicht behandelt:

- Programmiertechniken: Logikprogrammierung, Programmierung von parallelen Prozessen, . . .
- Programmierung von Web-Anwendungen, GUI-Programmierung, . . .
- Programmverifikation und Qualitätssicherung
- Werkzeuge
- Testen (unit tests (Modultest), system testing, regression testing)
- Modellierung (Flussdiagramme, Struktogramme, Klassendiagramme, unified modeling language (UML))

Wie geht es nun weiter?

Lesen Sie einmal Peter Norvigs Artikel "Teach yourself programming in ten years!" (`http://www.norvig.com/21-days.html`)

Er empfiehlt unter anderem:

- Finden Sie einen Weg, um Spaß am Programmieren zu finden! (Wenn nicht mit C++, dann mit einfacheren Sprachen wie Python oder Scheme.)
- Sprechen Sie mit anderen Programmierern, und lesen Sie anderer Leute Programme.
- Programmieren Sie selber!
- Wenn es Ihnen Spaß macht, besuchen Sie weitere Informatik-Vorlesungen.

- Arbeiten Sie an Projekten mit. Seien Sie einmal schlechtester, einmal bester Programmierer bei einem Projekt.

- Lernen Sie mehrere unterschiedliche Programmiersprachen.

- Lernen Sie Ihren Computer näher kennen (wie lang braucht eine Instruktion, ein Speicherzugriff, etc.)

Ich möchte noch folgende recht bodenständige Tips hinzufügen:

- Englisch ist heutzutage die Sprache sowohl der Naturwissenschaften als auch der Programmierung. Üben Sie Ihr Englisch, wo immer Sie können!

- Wenn Sie noch nicht mit zehn Fingern tippen können, versuchen Sie es zu lernen, sobald Sie einmal einen langen Text zu schreiben haben (z. B. Bachelor- oder Masterarbeit).