# Algorithmisches Denken und imperative Programmierung

11. Vorlesung

Janis Voigtländer

Universität Bonn

Wintersemester 2012/13

## Einschränkung von C zu einer Teilsprache C<sub>0</sub>

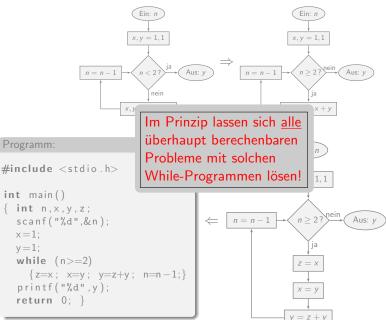
#### Motivation:

- ▶ Wir wollen C-Programmen eine formale Bedeutung geben (neben Hoare-Kalkül), ...
- ▶ außerdem allgemeine Prinzipien eines Compilers betrachten, . . .
- um die Dinge einfach zu halten, jedoch nicht alle Sprachkonstrukte behandeln.

#### Einschränkungen:

- nur int als Typ (keine Structs, keine Pointer, keine Arrays)
- keine Funktionsdefinitionen (außer main)
- keine globalen Variablendeklarationen, keine Konstantendeklarationen
- ► lediglich **while**-Schleifen und **if**-Verzweigungen als Kontrollstrukturen
- keine Kommentare; und bestimmte andere syntaktische Freiheitsgrade ebenfalls entfernt

# While-Programme — am Fibonacci-Beispiel



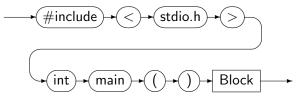
# Beispiel für ein C<sub>0</sub>-Programm

```
#include <stdio.h>
int main()
{ int i,n,s;
  scanf("%d",&n);
  i = 1:
  s=0:
  while (i \le n)
    \{ s=s+i*i;
     i = i + 1:
  printf("%d",s);
  return 0:
```

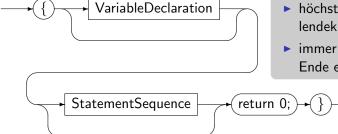
Formale Beschreibung: Syntaxdiagramme

# Syntaxdiagramme für C<sub>0</sub>: Programmstruktur (I)

#### **Program**



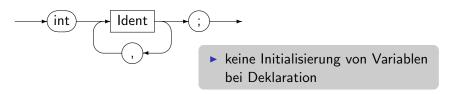
#### **Block**



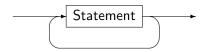
- höchstens eine Variablendeklarationszeile
- immer (und nur) am Ende ein return 0;

# Syntaxdiagramme für C<sub>0</sub>: Programmstruktur (II)

#### **VariableDeclaration**

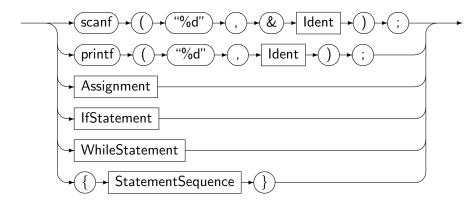


#### **StatementSequence**



# Syntaxdiagramme für C<sub>0</sub>: Programmstruktur (III)

#### Statement



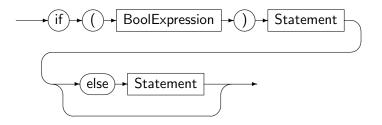
- bei printf kein beliebiger Ausdruck als auszugebendes Argument
- keine do-while-Schleife, keine for-Schleife

# Syntaxdiagramme für C<sub>0</sub>: Programmstruktur (IV)

#### **Assignment**



#### **IfStatement**



#### WhileStatement

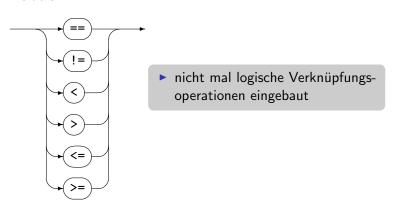


# Syntaxdiagramme für C<sub>0</sub>: Ausdrücke (I)

#### **BoolExpression**

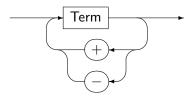


#### Relation

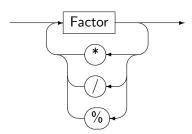


# Syntaxdiagramme für C<sub>0</sub>: Ausdrücke (II)

#### **SimpleExpression**

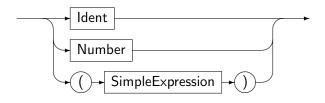


#### **Term**



# Syntaxdiagramme für C<sub>0</sub>: Ausdrücke (III)

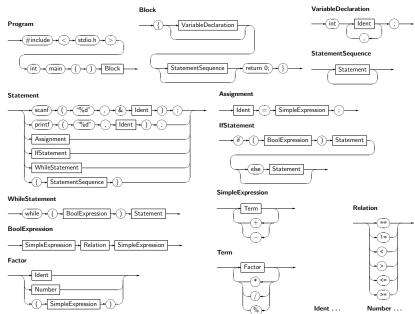
#### **Factor**



Ident Number

(lexikalische Syntax wie in C, natürlich nur ganze Zahlen)

# **Zusammenfassung Syntax C**<sub>0</sub>



## Kontextsensitive Nebenbedingungen in C<sub>0</sub>

Nicht durch Syntaxdiagramme ausdrückbare Forderungen an C-Programme gelten weiterhin, insbesondere:

- ▶ Jeder Bezeichner darf höchstens einmal in der eventuell vorhandenen VariableDeclaration deklariert sein.
- Wenn ein Bezeichner in der eventuell vorhandenen StatementSequence des Blocks auftritt, so muss er in der (dann unbedingt vorhandenen) VariableDeclaration des Blocks deklariert sein.

Insbesondere ist somit jedes  $C_0$ -Programm auch ein gültiges C-Programm.

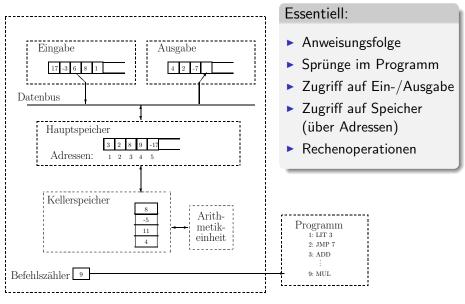
Realisierung in "Maschinensprache"

# Modellierung/Umsetzung von C<sub>0</sub>-Sprachfeatures

```
#include <stdio.h>
int main()
{ int i,n,s;
  scanf("%d",&n);
  i = 1;
  s=0:
  while (i \le n)
    \{ s=s+i*i;
     i=i+1:
  printf("%d",s);
  return 0;
```

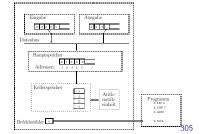
Wir brauchen: Ein- und Ausgabe, Speicherplätze, Auswertung von Ausdrücken, Zuweisungen, Kontrollfluss

# **Eine abstrakte Maschine (AM)**



## Modellierung des Hauptspeichers

- statt benannter Speicherplätze (Variablen) lediglich numerische Adressen
- einzig möglicher Speicherinhalt: ganze Zahl (int)
- konkreter Hauptspeicherinhalt modelliert als partielle Abbildung von Adressen auf Inhalte
- ▶ mathematisch:  $HS = \{h \mid h : \mathbb{N}_+ \longrightarrow \mathbb{Z}\}$
- ▶ Notation: etwa h = [1/3, 2/2, 3/5]
- ▶ Update: h[n/d]; h[n/d](n') = d, falls n' = n, sonst h[n/d](n') = h(n')
- ▶ Beispiele für h = [1/3, 2/2, 3/5]: h[5/7] = [1/3, 2/2, 3/5, 5/7] und h[2/7] = [1/3, 2/7, 3/5]
- Befehle zur Kommunikation mit:
  - ► Ein- und Ausgabe
  - "Berechnungseinheit"

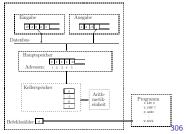


### Modellierung von Ein- und Ausgabe

- ▶ lediglich Ein- und Ausgabe ganzer Zahlen (int)
- ► Abstraktion von interaktiver Ein- und Ausgabe; stattdessen: Eingabeband und Ausgabeband
- ▶ Bandinhalte modelliert als endliche Listen
- ightharpoonup mathematisch: Inp = Out =  $\mathbb{Z}^*$
- ▶ Notation: etwa inp = 1.13.5 oder  $out = \varepsilon$
- ► READ *n*: Lesen/Konsumieren erster Position von *inp* und entsprechendes Update der Position *n* des HS

▶ WRITE n: Ausgabe des HS-Inhalts von Position n

am Ende von out



## Auswertung von Ausdrücken, Zuweisungen

#### Problem:

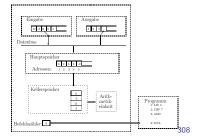
- ► Wir wollen strukturierte Ausdrücke wie (3+5)\*(7-2) oder j>2\*i+1 auswerten.
- Dies soll jedoch in "linearisierter Form" erfolgen, insbesondere soll eine Operation immer nur auf zwei Operanden wirken.
- Folglich müssen Zwischenergebnisse berechnet und in geeigneter Weise vorrätig gehalten werden.

#### Lösung:

- ► Verwendung eines "Datenkellers"
- Ablage und Entnahme jeweils nur an Kellerspitze
- ▶ mathematisch: DK = Z\*
- ► Notation: etwa *d* = 8:7:2
- ▶ LIT z: Ablage einer Konstante
- ▶ LOAD *n*: Ablage des HS-Inhalts von Position *n*
- ► ADD, MUL, ..., LT, EQ, NE, ...: Berechnungen
- ► STORE *n*: Entnahme und entsprechendes Update der Position *n* des HS

## Kontrollfluss — Abarbeitungsreihenfolge

- ▶ Durchnummerierung aller Befehle in einem Programm
- ▶ aktuelle Position im Befehlszähler gespeichert ( $m \in BZ = \mathbb{N}$ )
- normalerweise einfache Erhöhung nach jeder Befehlsabarbeitung
- ▶ JMP *n*: Sprung an Position *n*
- ▶ JMC n: bedingter Sprung an Position n; wenn Spitze des Datenkellers gleich 0, dann Sprung; wenn Spitze des Datenkellers gleich 1, dann weiter mit nächstem Befehl



## Also, vollständige Befehlsübersicht:

- ▶ READ n: Lesen von Eingabeband in Hauptspeicher
- ▶ WRITE n: Ausgabe aus Hauptspeicher auf Ausgabeband
- ▶ LOAD *n*: Ablage aus Hauptspeicher auf Datenkeller
- STORE n: Entnahme aus Datenkeller in Hauptspeicher
- ▶ LIT z: Ablage einer Konstante auf Datenkeller
- ► ADD, MUL, SUB, DIV, MOD, LT, EQ, NE, GT, LE, GE: Berechnungen und Vergleiche (auf Datenkeller)
- ▶ JMP *n*: Sprung
- ▶ JMC *n*: bedingter Sprung abhängig von Wert auf Datenkeller

### Formale Modellierung

▶  $AM = BZ \times DK \times HS \times Inp \times Out$ , mit:

```
\begin{array}{lll} \mathsf{BZ} &= \mathbb{N} & \mathsf{Befehlsz\ddot{a}hler} \\ \mathsf{DK} &= \mathbb{Z}^* & \mathsf{Datenkeller} \\ \mathsf{HS} &= \{ \textit{h} \mid \textit{h} : \mathbb{N}_+ - \!\!\!\! \rightarrow \mathbb{Z} \} & \mathsf{Hauptspeicher} \\ \mathsf{Inp} &= \mathbb{Z}^* & \mathsf{Eingabeband} \\ \mathsf{Out} &= \mathbb{Z}^* & \mathsf{Ausgabeband} \end{array}
```

- ▶ Maschine ist jederzeit in einem Zustand  $s = (m, d, h, inp, out) \in AM$ .
- ▶ Ein Programm ist eine partielle, endlich definierte Abbildung P von  $\mathbb{N}$  auf die Menge  $\Gamma$  aller Befehle.
- Notation: 1: P(1); 2: P(2); 3: P(3);
- Programmabarbeitung besteht aus wiederholter Anwendung von Befehlen auf einen Zustand.

# Befehlssemantik (I)

```
C[\![.]\!]:\Gamma\longrightarrow (AM-\rightarrow AM)
C[READ n](m, d, h, inp, out) =
   wenn inp = first(inp).rest(inp) mit first(inp) \in \mathbb{Z}, rest(inp) \in \mathbb{Z}^*,
   dann (m+1, d, h[n/first(inp)], rest(inp), out)
C[WRITE n](m, d, h, inp, out) =
    wenn h(n) \in \mathbb{Z}, dann (m+1, d, h, inp, out.h(n))
C[LOAD n](m, d, h, inp, out) =
    wenn h(n) \in \mathbb{Z}, dann (m+1, h(n) : d, h, inp, out)
C[STORE n](m, d, h, inp, out) =
    wenn d = d.1 : d', dann (m + 1, d', h[n/d.1], inp, out)
C[[L]Tz](m,d,h,inp,out) = (m+1,z:d,h,inp,out)
```

# Befehlssemantik (II)

```
C[ADD](m, d, h, inp, out) =
   wenn d = d.1 : d.2 : d', dann (m + 1, (d.2 + d.1) : d', h, inp, out)
für MUL, SUB, DIV und MOD analog
C[LT](m, d, h, inp, out) =
   wenn d = d.1 : d.2 : d', dann (m + 1, b : d', h, inp, out),
   wobei b = 1, falls d.2 < d.1, sonst b = 0
für EQ, NE, GT, LE und GE analog
C[JMP e](m, d, h, inp, out) = (e, d, h, inp, out)
C[JMC e](m, d, h, inp, out) =
   wenn d = 0: d', dann (e, d', h, inp, out);
   wenn d = 1: d', dann (m + 1, d', h, inp, out)
```

## Programmsemantik — am Beispiel

```
15: STORE 3:
 1: READ 2:
                 8: LE:
 2: LIT 1:
         9: JMC 21;
                                16: LOAD 1:
 3: STORE 1;
                10: LOAD 3; 17: LIT 1;
 4: LIT 0;
                11: LOAD 1; 18: ADD;
 5: STORE 3;
                12: LOAD 1; 19: STORE 1;
 6: LOAD 1;
                13: MUL; 20: JMP 6;
 7: LOAD 2;
                14: ADD; 21: WRITE 3;
C[READ n](m, d, h, inp, out) =
  wenn inp = first(inp).rest(inp) mit first(inp) \in \mathbb{Z}, rest(inp) \in \mathbb{Z}^*,
  dann (m+1, d, h[n/first(inp)], rest(inp), out)
n=2, m=1, d=\varepsilon, h=[], inp=2, out =\varepsilon
```

# Programmsemantik — am Beispiel

```
15: STORE 3:
 1: READ 2; 8: LE;
 2: LIT 1:
           9: JMC 21; 16: LOAD 1;
 3: STORE 1;
                     10: LOAD 3; 17: LIT 1;
 4: LIT 0;
                     11: LOAD 1; 18: ADD;
 5: STORE 3; 12: LOAD 1; 19: STORE 1;
 6: LOAD 1; 13: MUL; 20: JMP 6;
 7: LOAD 2: 14: ADD; 21: WRITE 3;
( 7 , 3 , [1/3,2/2,3/5] , \varepsilon , \varepsilon ) ( 8 , 2:3 , [1/3,2/2,3/5] , \varepsilon , \varepsilon )
( 9, 0, [1/3,2/2,3/5], \varepsilon, \varepsilon)
( \frac{21}{2} , \varepsilon , [1/3,2/2,3/5] , \varepsilon , \varepsilon )
(22. \varepsilon, [1/3,2/2,3/5], \varepsilon, 5)
C[WRITE n](m, d, h, inp, out) =
   wenn h(n) \in \mathbb{Z}, dann (m+1, d, h, inp, out.h(n))
```

### Programmsemantik — formal

(zur Erinnerung: Befehlssemantik  $\mathcal{C}[\![.]\!]:\Gamma\longrightarrow (\mathsf{AM}\longrightarrow \mathsf{AM})$ )

Sei Prog die Menge aller AM-Programme.

$$\mathcal{I}\llbracket.\rrbracket : Prog \longrightarrow (\mathsf{AM} \longrightarrow \mathsf{AM})$$

$$\mathcal{I}\llbracketP\rrbracket(m,d,h,inp,out) = \begin{cases} \mathcal{I}\llbracketP\rrbracket(\mathcal{C}\llbracket\mathsf{P}(\mathsf{m})\rrbracket(m,d,h,inp,out)), & \text{falls } m \in def(\mathsf{P}) \\ (m,d,h,inp,out), & \text{falls } m \notin def(\mathsf{P}) \end{cases}$$

$$\mathcal{P}\llbracket.\rrbracket : Prog \longrightarrow (\mathsf{Inp} \longrightarrow \mathsf{Out})$$

$$\mathcal{P}\llbracket\mathsf{P}\rrbracket(inp) = proj_5^{(5)}(\mathcal{I}\llbracket\mathsf{P}\rrbracket(1,\varepsilon,\llbracket],inp,\varepsilon))$$

▶ proj<sub>5</sub><sup>(5)</sup> meint Projektion auf fünfte Komponente eines Fünftupels.

```
    Ausgabe der Gesamtsumme auf Ausgabeband

1: LIT 0;
           6: NE; 11: STORE 2;
2: STORE 2; 7: JMC 13; 12: JMP 3:
3: READ 1; 8: LOAD 2; 13: WRITE 2;
4: LOAD 1; 9: LOAD 1;
5: LIT 0: 10: ADD:
```

```
    Ausgabe der Gesamtsumme auf Ausgabeband

 1: LIT 0:
                   6: NE: 11: STORE 2:
 2: STORE 2; 7: JMC 13; 12: JMP 3:
 3: READ 1: 8: LOAD 2: 13: WRITE 2:
 4: LOAD 1: 9: LOAD 1:
 5: LIT 0:
                  10: ADD:
( 8, \varepsilon, [1/3,2/0], 4.2.0 , \varepsilon)
(9, 0, [1/3,2/0], 4.2.0, \varepsilon)
(10, 3:0, [1/3,2/0], 4.2.0, \varepsilon)
(11, 3, [1/3,2/0], 4.2.0, \varepsilon)
(12, \varepsilon, [1/3,2/3], 4.2.0 , \varepsilon)
```

```
    Ausgabe der Gesamtsumme auf Ausgabeband

 1: LIT 0:
                     6: NE: 11: STORE 2:
 2: STORE 2; 7: JMC 13; 12: JMP 3:
 3: READ 1: 8: LOAD 2: 13: WRITE 2:
 4: LOAD 1: 9: LOAD 1:
 5: LIT 0:
                    10: ADD:
(5, 4, [1/4,2/3], 2.0
                            , \varepsilon
(6, 0.4, [1/4,2/3], 2.0, \varepsilon)
(7, 1, [1/4,2/3], 2.0
                            , \varepsilon)
( 8 , \varepsilon , [1/4,2/3] , 2.0
                            , \varepsilon)
(9, 3, [1/4,2/3], 2.0
                             , \varepsilon)
(10, 4:3, [1/4,2/3], 2.0
                             , \varepsilon
(11, 7, [1/4,2/3], 2.0
                             , \varepsilon
(12, \varepsilon, [1/4,2/7], 2.0
                             . \varepsilon)
```

```
    Ausgabe der Gesamtsumme auf Ausgabeband

 1: LIT 0:
                6: NE: 11: STORE 2:
 2: STORE 2; 7: JMC 13; 12: JMP 3;
 3: READ 1: 8: LOAD 2: 13: WRITE 2:
 4: LOAD 1: 9: LOAD 1:
 5: LIT 0:
               10: ADD:
(12, \varepsilon, [1/4,2/7], 2.0 , \varepsilon)
( 3, \varepsilon, [1/4,2/7], 2.0 , \varepsilon)
( 6, 0:2, [1/2,2/7], 0
                      , \varepsilon)
```

```
    Ausgabe der Gesamtsumme auf Ausgabeband

  1: LIT 0:
                               6: NE: 11: STORE 2:
 2: STORE 2; 7: JMC 13; 12: JMP 3:
 3: READ 1; 8: LOAD 2; 13: WRITE 2;
  4: LOAD 1: 9: LOAD 1:
  5: LIT 0:
                             10: ADD:
(9, 7, [1/2,2/7], 0
                                         , \varepsilon
                                         , \varepsilon
(10, 2:7, [1/2,2/7], 0
(11, 9, [1/2,2/7], 0
                                         , \varepsilon
( 12 , \varepsilon , [1/2,2/9] , 0
                                         , \varepsilon
( 3, \varepsilon, [1/2,2/9], 0
                                         , \varepsilon)
 \left( \begin{array}{ccc} 4 \;, & \varepsilon \;, \, [1/0,2/9] \;, \; \varepsilon & , \; \varepsilon \; \right) \\ \left( \begin{array}{ccc} 5 \;, & 0 \;, \, [1/0,2/9] \;, \; \varepsilon & , \; \varepsilon \; \right) \end{array} 
(6.0:0.[1/0.2/9], \varepsilon
                                         , \varepsilon
```

Aufgabe:

```
    Ausgabe der Gesamtsumme auf Ausgabeband

 1: LIT 0:
                        6: NE: 11: STORE 2:
 2: STORE 2; 7: JMC 13; 12: JMP 3;
 3: READ 1: 8: LOAD 2: 13: WRITE 2:
 4: LOAD 1: 9: LOAD 1:
 5: LIT 0:
                       10: ADD:
                                , \varepsilon)
(12, \varepsilon, [1/2,2/9], 0
(3, \varepsilon, [1/2,2/9], 0
                                , \varepsilon)
                                          Bzgl. Speicherplatz 2
( 4, arepsilon, [1/0,2/9], arepsilon
                                , \varepsilon)
                                          zeigt sich hier ein sehr
(5, 0, [1/0,2/9], \varepsilon
                                , \varepsilon)
                                          spezielles Zugriffsschema,
( 6, 0:0, [1/0,2/9], \varepsilon
                                , \varepsilon)
                                          welches man ausnutzen
(7, 0, [1/0,2/9], \varepsilon
                                          kann, um das Programm
                                , \varepsilon
(13, \varepsilon, [1/0,2/9], \varepsilon
                                          zu optimieren.
                                , \varepsilon
(14, \varepsilon, [1/0,2/9], \varepsilon
                                 .9)
```

Einlesen einer Folge (Ende: 0) vom Eingabeband

#### **Optimierung**

Idee: statt Ablage im HS, Zwischensumme auf DK lassen

```
1: LIT 0; 5: NE; 9: JMP 2; 2: READ 1; 6: JMC 10; 10: STORE 1; 3: LOAD 1; 7: LOAD 1; 11: WRITE 1; 4: LIT 0; 8: ADD;  ( \ 1 \ , \ \varepsilon \ , \ [] \ , \ 3.4.2.0 \ , \varepsilon )
```

### **Optimierung**

Idee: statt Ablage im HS, Zwischensumme auf DK lassen

```
1: LIT 0: 5: NE: 9: JMP 2:
 2: READ 1; 6: JMC 10; 10: STORE 1;
 3: LOAD 1; 7: LOAD 1; 11: WRITE 1;
 4: LIT 0; 8: ADD;
 8 , 2:7 , [1/2] , 0 , arepsilon )
 9 , 9 , [1/2] , 0 , arepsilon )
 2 , 9 , [1/2] , 0 , \varepsilon )
(3, 9, [1/0], \varepsilon
(4, 0:9, [1/0], \varepsilon , \varepsilon)
( 5, 0:0:9, [1/0], \varepsilon
                            , \varepsilon)
( 6, 0:9, [1/0], \varepsilon
                            , \varepsilon)
(10, 9, [1/0], \varepsilon , \varepsilon)
(11, \varepsilon, [1/9], \varepsilon , \varepsilon)
( 12 , \varepsilon , [1/9] , \varepsilon
                            , 9)
```

### **Erweiterung**

#### Aufgabe: statt Summe, nun Durchschnitt zu berechnen

```
1: LIT 0:
                8: JMC 16: 15: JMP 4:
2: STORE 2; 9: LOAD 1; 16: LOAD 2;
3: LIT 0:
              10: ADD:
                            17: DIV:
              11: LOAD 2; 18: STORE 1;
4: READ 1;
5: LOAD 1;
              12: LIT 1:
                           19: WRITE 1:
6: LIT 0:
               13: ADD:
7: NE:
               14: STORE 2:
(1, \varepsilon, []
                , 3.4.2.0 , \varepsilon )
```

### **Erweiterung**

#### Aufgabe: statt Summe, nun Durchschnitt zu berechnen

```
1: LIT 0:
                 8: JMC 16:
                                 15: JMP 4:
2: STORE 2;
                 9: LOAD 1:
                                 16: LOAD 2:
3: LIT 0:
                10: ADD:
                                 17: DIV:
                11: LOAD 2; 18: STORE 1;
4: READ 1:
5: LOAD 1:
                12: LIT 1:
                                 19: WRITE 1:
6: LIT 0:
                13: ADD:
7: NE:
                14: STORE 2:
```

#### Programmieren in AM:

- ist ja ganz nett
- erlaubt Ausnutzen von in C/C<sub>0</sub> gar nicht sichtbaren Details der Maschine
- ist aber auch umständlich (z.B. keine Strukturierung)

Daher: Übersetzung  $C_0 \rightarrow AM$