Praktikum Rechnerstrukturen

4

Assemblerebene
Stack

Name, Vorname	
Bogen bearbeitet	abzeichnen lassen

Universität Hamburg – MIN – Fachbereich Informatik
Arbeitsbereich Technische Aspekte Multimodaler Systeme

https://tams.informatik.uni-hamburg.de

1 Die Assemblerebene

Inhalt dieses Versuchs ist die Programmierung auf der Assemblerebene. Schon beim Erstellen der kurzen Maschinenprogramme dürfte klargeworden sein, dass die Programmierung in der reinen Maschinensprache sowohl extrem (zeit-) aufwändig als auch fehleranfällig ist. Ohne weitere Unterstützung lassen sich auf diese Weise kaum Programme mit mehr als einigen hundert Befehlen sinnvoll erstellen.

Andererseits haben Sie beim Erstellen der Maschinenprogramme bereits alle Funktionen kennen gelernt, die ein Assembler automatisiert — vom Zusammensetzen von Opcode und Registerangaben zu vollständigen Befehlsworten bis zur Berechnung von Sprungadressen.

Merkmal einer *Assemblersprache* ist die 1:1-Abbildung jedes Assemblerbefehls auf einen Maschinenbefehl. Wegen dieser direkten Zuordnung zu Maschinenbefehlen sind Assembler und ihre Eingabesprachen normalerweise auf eine bestimmte Architektur zugeschnitten. Es gibt aber auch universelle Assembler (wie der Assembler des GNU-Projektes *GAS* oder der *vasm*) die für eine Reihe von verschiedenen Architekturen und Prozessoren benutzt werden können. Wichtige gemeinsame Merkmale aller Assemblersprachen sind die folgenden:

- Verwendung von einprägsamen Namen für die einzelnen Befehle (*Mnemonics*)
- einfache und reguläre Syntax für Befehlsargumente wie Register oder Speicheradressen
- Definition von symbolischen Namen für Konstanten und Sprungmarken
- Unterstützung von Kommentaren und freie Formatierung
- Umrechnung der symbolischen Programmadressen in die physikalischen Adressen
- Erstellen von Hilfsdateien, etwa eine Liste der verwendeten Namen, Sprungmarken usw.
- voller Zugriff auf alle Befehle und Register des benutzten Prozessors
- evtl. Unterstützung fortgeschrittener Techniken, etwa das Einbinden mehrerer Quelldateien mittels `include` oder Makrofähigkeit
- Häufig wird der eigentliche Assembler um weitere Tools wie Debugger und Disassembler ergänzt. Damit können Details völlig vom Benutzer ferngehalten werden (etwa die Umrechnung zwischen Byte- und Wortadressen).

Obwohl ein Assemblerprogramm weiterhin auf der Ebene einzelner Befehle geschrieben wird, ist der Produktivitätsgewinn gegenüber der Maschinensprache beträchtlich. Auf der anderen Seite ist die Assemblerprogrammierung natürlich immer noch sehr viel aufwändiger als die Programmierung in Hochsprachen (wie Java oder C usw.). Trotzdem gibt es Gründe, in Assembler zu programmieren:

- es steht (noch) kein geeigneter Compiler für eine Hochsprache zur Verfügung
- kritische Programmanteile erfordern maximale Performance
- Zugriff auf Spezialregister und privilegierte Register, etwa für Gerätetreiber
- eingeschränkte Resourcen an Programm- und Datenspeicher — viele 8-bit Mikrocontroller enthalten weniger als 1KByte RAM

1.1 Format der Assemblersprache

Obwohl jede Assemblersprache auf die Struktur der Befehle der zugrundeliegenden Architektur zugeschnitten ist, ähneln sich die Assemblersprachen für verschiedene Prozessoren doch sehr stark. Fast immer werden die Programme mit genau einer Assemberanweisung pro Zeile geschrieben, und jede Zeile wiederum beginnt mit einer optionalen Marke, gefolgt vom Befehl (Opcode), den Operanden und einem optionalen Kommentar. Der Assembler für den D·CORE (`winT3asm.exe` Version für Windows oder die plattformunabhängige Java-Version `t3asm.jar`) verwendet das folgende Format für die Eingabedateien:

```
; strtoint.asm
; Umwandeln eines nullterminierten Strings in eine Zahl:
; der String steht ab Adresse 0x8000 im Speicher und
; das Ergebnis im Register R12.

Start:
          movi    r10, 8          ; R10 = 8
          lsli    r10, 12         ; R10 = 0x8000 Stringadresse
          movi    r0, 0           ; zum Vergleich
          movi    r12, 0          ; Zahl initialisieren
Schleife:
          ldw     r1, 0(r10)      ; Character laden
          cmpe    r1, r0          ; = 0?  (Ende des Strings)
          bt      ende
          andi    r1, 0xf         ; Character -> Zahl
          addu    r12, r12        ; 2*r12_alt
          mov     r2, r12         ; Sichern
          lsli    r12, 2          ; 4*r12= (8* r12_alt)
          addu    r12, r2         ; 10*R12_alt
          addu    r12, r1         ; + Zahl
          addi    r10, 2          ; Adresse erhöhen
          br      Schleife
ende:
          halt

         .org     0x8000          ; Adresszähler auf 0x8000
         .ascii   "1324"
         .defw  0                 ; Null-Wort als String-Ende
         .end                     ; Kann auch weggelassen werden
```

Die Details finden Sie in der ausführlichen, separaten Beschreibung `t3asm.pdf` für den Assembler. Zusammengefasst gelten die folgenden Regeln für das Eingabeformat:

- *Kommentare* beginnen mit `;` und reichen bis zum Zeilenende.
- *Label-Definitionen* sind Strings, die in der ersten Spalte der Datei beginnen und mit einem Doppelpunkt abgeschlossen werden.
- *Hex-Konstanten* werden in der Schreibweise `0xCAFE` erwartet.
- Die *.org-Direktive* sorgt dafür, dass die nachfolgenden Befehle oder Konstanten ab der angegebenen Adresse ⟨*addr*⟩ im ROM/RAM abgelegt werden.
- Die *.defw-Direktive* dient dazu, ein bestimmtes Datenwort in die jeweilige Speicherstelle zu schreiben.

- Die *.defs-Direktive* reserviert die angegebene Anzahl von Speicherworten.
- Die *.ascii-Direktive* erlaubt es, Zeichenketten im ROM/RAM abzulegen, mit jeweils einem ASCII-Zeichen pro Speicherwort.

Aufgabe 4.1 Unterprogramme

Vergegenwärtigen Sie sich noch einmal die Arbeitsweise des Befehles `JSR` (*Jump to Subroutine*), insbesondere wie und wohin aus dem Unterprogramm zurückgekehrt wird.

Was bewirkt das folgende Unterprogramm *LDR5*?

```
LDR5:   ldw     r5, 0(r15)
        addi    r15, 2
        JMP     r15
```

Aufgerufen wird es z.B. durch die Sequenz:

```
        ⋮
        jsr     LDR5
        .defw   0xAFFE   ; oder dezimal .defw 45054
        ⋮
```

Notieren Sie dazu die Registerinhalte nach Ausführung der folgenden Befehle (beim `JSR` ist `0x200` bereits die Adresse zu der gesprungen werden soll und **nicht** der dazu erforderliche Offset):

Adresse	Befehl/Daten	PC	R0	R5	R15
0x009F	.....	0x0100	0xFFFF	0x1234	0x369C
0x0100	movi r0, 0				
0x0102	jsr 0x0200				
0x0104	.defw 0x8010				
0x0106	ldw r0, 0(r5)				
.....	.....				
0x0200	ldw r5, 0(r15)				
0x0202	addi r15, 2				
0x0204	jmp r15				
.....	.....				
0x8010	.defw 0xAFFE				

Hinweis: Bei allen noch folgenden Aufgaben geben Sie am besten Ihren Code mit dem integrierten Editor des Assemblers/Emulators ein, um ihn dann zu übersetzen und auszuführen. Ggf. benötigte „Templates“ sind in `t3-hades/programs` hinterlegt, dort sollten Sie auch eigene Programme abspeichern!

Den auf den Praktikumsrechnern installierten D·CORE-Assembler/Emulator starten Sie mit dem Shell-Kommando: **`t3asm`**.
Für Arbeiten zu Hause können Sie die plattformunabhängige Java-Version des D·CORE-Assemblers (*T3asm.jar*) verwenden. Unter Linux starten sie ihn beispielsweise mit dem Shell-Kommando: `java -jar T3asm.jar`.

2 Maschinenarithmetik

Die nächsten beiden Aufgaben dienen dazu, noch einmal einige Aspekte der Zahldarstellung und der Integerarithmetik aufzufrischen.

Aufgabe 4.2 Quadrate

In Bogen 2, Aufgabe **??** und Bogen 3, Aufgabe **??** hatten wir eine einfache Möglichkeit betrachtet, das Quadrat einer Zahl $n \geq 0$ ohne Multiplikationen zu berechnen.

Schreiben Sie jetzt ein Hauptprogramm, das das Unterprogramm *nh2* aus der Datei `Quadrat.asm` aufruft und berechnen Sie mit seiner Hilfe das Quadrat der Zahlen 5, 55 und 101. *nh2* erwartet im Register `R5` die Zahl n, deren Quadrat berechnet werden soll, und liefert im Register `R7` das Ergebnis zurück. Weiterhin zerstört das Unterprogramm den Inhalt des Registers `R8`.

Label	Adresse	Befehl (Mnemonic, Operanden)	Kommentar
	0000		
	0002		
	0004		
	0006		
	0008		
	000a		
	000c		
	000e		
	0010		
	0012		
	0014		
	0016		
	0018		
	001a		
	001c		
	001e		

Wie groß darf n für ein richtiges Ergebnis höchstens sein, wenn

(a) der Wert im Register `R7` als Integer interpretiert wird:

(b) der Wert im Register `R7` als Unsigned interpretiert wird:

Aufgabe 4.3 Quadrate II

Erweitern Sie das Unterprogramm *nh2*, so dass auch größere Zahlen als Eingabe für ***n*** verarbeitet werden können. Das möglicherweise 32-bit große Ergebnis passt dann natürlich nicht mehr in nur ein Register, sondern es werden zwei Register (z.B. `R6` für die hochwertigen 16-bit und `R7` für die niederwertigen 16-bit) benötigt.
Gehen Sie der Einfachheit halber von einer unsigned Arithmetik aus. Testen Sie Ihr Programm mit $n = 256$ und $n = 1234$.

Label	Adresse	Befehl (Mnemonic, Operanden)	Kommentar

Aufgabe bearbeitet	abzeichnen lassen

3 Adressierungsarten und Zeichenketten

Die Befehle des Prozessors verwenden verschiedene *Adressierungsarten*. Die wichtigsten sind (angepasst auf unseren D·CORE):

unmittelbare Adressierung
: Im Befehlswort steht ein **Zahlwert** *W* und ein Register *REG*. Aus dem Inhalt von *REG* und der Zahl *W* wird ein neuer Wert berechnet und nach *REG* geschrieben. Es ist auch der Fall denkbar, dass *REG* nicht explizit angegeben wird, weil es sich aus der besonderen Art des Befehls ergibt.

direkte Adressierung
: Im Befehlswort steht die **absolute Adresse** *ADR* einer Speicherstelle in ROM/RAM und ein Register *REG*.

Registeradressierung
: Im Befehlswort stehen normalerweise **zwei Register**, deren Inhalt verknüpft und dann in eins der beiden Register geschrieben wird. Es ist aber auch der Fall denkbar, dass nur eines der Register explizit angegeben wird, und das andere sich aus der besonderen Art des Befehls ergibt. Im Extremfall kann sogar auch die Angabe dieses Registers fehlen.

indirekte Registeradressierung
: Im Befehlswort stehen **zwei Register** *REG*1 und *REG*2. Der Inhalt von *REG*1 wird als Adresse einer Speicherstelle in ROM/RAM interpretiert, in die der Inhalt von *REG*2 geschrieben wird, bzw. deren Inhalt nach *REG*2 gebracht wird.

indizierte Adressierung
: Wie die indirekte Adressierung, nur steht im Befehlswort zusätzlich noch ein Wert, der auf die ROM/RAM-Adresse addiert wird.

Aufgabe 4.4 Adressierungsarten

Welche der Adressierungsarten werden beim D·CORE verwendet? Beachten Sie, dass auch der *PC* ein Register ist, obwohl er nicht in der Registerbank liegt.

(a) für die arithmetischen Befehle (z.B MOV und MOVI):

(b) für Speicherzugriffe (LDW und STW):

(c) für den JMP-Befehl:

(d) für die Branch-Befehle (BR, BT und BF):

(e) für den JSR-Befehl:

Eine besonders wichtige Anwendung von Arrays und indizierter Adressierung sind Zeichenketten (Strings). In C und verwandten Sprachen wird eine Zeichenkette nur durch ihre Speicheradresse spezifiziert. Alle nachfolgenden Bytes bis zum ersten Null-Byte `0x00` (einschließlich) stellen die Zeichenkette dar. Einige andere Sprachen benutzen statt dessen eine zusammengesetzte Datenstruktur mit einem Integer für die Anzahl der Zeichen und einem separaten Array für die einzelnen Zeichen.

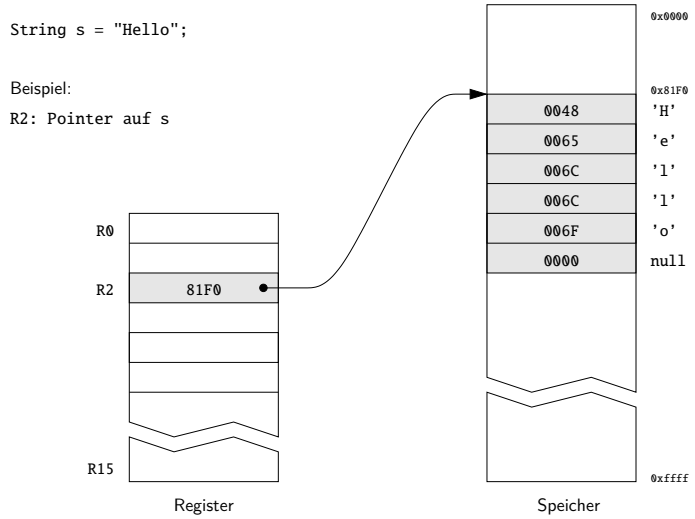


Abb. 1: Null-terminierter String im Speicher, ein Zeichen pro Wort

Aufgabe 4.5 strlen()

Erstellen Sie ein Assemblerunterprogramm zusammen mit einem aufrufenden Hauptprogramm für die Funktion `strlen()`, das die Länge einer Zeichenkette (ohne das terminierende Nullbyte!) zurückliefert. Die Startadresse des Strings stehe in `R10`, das Resultat (die Länge des Strings) soll in `R11` zurückgeliefert werden. Verwenden Sie die C-Konvention mit null-terminierten Strings und 16-Bit pro Zeichen, wie in Abbildung 4 illustriert.

Einen String bekommen Sie dabei mit der Befehlsfolge

```
.org   0x8000           ; Adresse
.ascii "Ein String"     ; der String
.defw  0                ; terminierende Null
```

ab der Adresse `0x8000` in den Speicher. Setzen Sie Sie diese Anweisungen bitte immer ganz an das Ende ihres Programms! Weshalb ist es wichtig, bzw. was wäre die Konsequenz wenn nicht?

Einen Rumpf finden Sie in der Datei `m_strlen.asm`, die Sie sich in den Assembler laden können.

Aufgabe bearbeitet	abzeichnen lassen

Aufgabe 4.6 num2str()

In der Datei `num2str.asm` finden Sie ein Assemblerprogramm, das den Inhalt eines Registers für die Ausgabe auf einem alpahnumerischen Terminal im HEX-Format vorbereitet. Hierfür ist eine Funktion `char* num2str(unsigned num, char* str)` erforderlich. Diese Funktion transformiert den Inhalt eines Registers, im vorliegenden Falle einen 16-bit Wert, in eine Stringdarstellung. Erstellen Sie das Assemblerunterprogram `num2str`, das für die Konvertierung benötigt wird.
Mit dem Pseudobefehl `.prstr` ⟨*reg*⟩ – die Anfangsadresse des Strings stehe im Register ⟨*reg*⟩ – lässt sich der String im Anzeigefeld des Emulators ausgeben.

Mit welcher Byte-Order ist die Reihenfolge der Speicherung der Zeichen im String (Endianness) vergleichbar?

- Little-Endian oder
- Big-Endian

Aufgabe bearbeitet	abzeichnen lassen

Hinweis zu obiger Pseudoanweisung

Die obige Pseudoanweisung gehört zu einer Klasse von Pseudoanweisungen, die normalerweise keinen Maschinencode erzeugen, sondern der Steuerung des Assemblers bzw. Linkers dienen. Der Pseudobefehl `.prstr` ⟨*reg*⟩ allerdings dient der Steuerung des Emulators und wird deshalb vom Assembler in einen Maschinenbefehl mit Opcode `0x0` umgesetzt (`0x0xxx`). Befehle mit diesem Opcode stellen ausschließlich Anweisungen an den Emulatior dar, der dann z. B. die gewünschten Ausgaben im Emulatorfenster erzeugt.
Bei Ausführung durch den D·CORE-Prozessor würde der Befehl mit dem Opcode `0x0` mit Standardmikroprogramm aus der Decode-Phase direkt in die Fetch-Phase des nächsten Befehls verzweigen; würde also ein `NOOP` (no operation) bewirken.

4 Speicherbereiche und Stack

Da beim von-Neumann Rechner sowohl die Programme als auch alle Daten im Hauptspeicher liegen, ist die Organisation des Speichers von zentraler Bedeutung. Die in Unix übliche Konvention zur Einteilung der Speicherbereiche ist in Abbildung 2 gezeigt. Dabei werden die folgenden Speicherbereiche (*Segmente*) unterschieden:

- Das *Textsegment* enthält den eigentlichen Programmtext mit allen Befehlen. Sofern keine selbstmodifizierende Programme zum Einsatz kommen, bleibt das Textsegment während des Programmablaufs unverändert. Es wird häufig am unteren Ende des Speichers abgelegt.
- Der *constant pool* (Konstantenbereich) nimmt alle Konstanten und statischen Variablen des Programms auf. Typ und Anzahl dieser Variablen ergeben sich unmittelbar aus dem Programm. Der Speicherplatz für diese Variablen wird normalerweise direkt oberhalb des *Textsegments* angelegt.
- Der *Heap* (Halde) nimmt alle dynamisch zur Laufzeit des Programms erzeugten Variablen bzw. Objekte auf. Der Heap wird oberhalb des Konstantenpools angelegt und wächst nach oben. Für den Heap werden (Betriebssystem-) Funktionen benötigt, um freie Speicherbereiche für neu anzulegende Variablen zu finden und diese auch wieder freigeben zu können.
- Der *Stack* (Stapel) wird für die Parameterübergabe zwischen Funktionen und für die Speicherung der lokalen Variablen der einzelnen Funktionen benutzt. Der Stack wird häufig ab dem oberen Ende des zur Verfügung stehenden Speichers angelegt und wächst mit jedem Aufruf nach unten.

Der Befehl `JSR` des D·CORE speichert die Rücksprungadresse in Register `R15`. Das bedeutet, dass ohne weitere Maßnahmen höchstens ein Unterprogramm aufgerufen werden kann, da sonst ein zweiter Aufruf die Rücksprungadresse des ersten Aufrufs überschreibt. Für geschachtelte Aufrufe muss daher ein *Stack* bereitgestellt und vom Anwenderprogramm aus verwaltet werden.

Wie bei fast allen RISC-Prozessoren (außer SPARC), gibt es im Befehlssatz keine weitere Unterstützung für die Stack-Verwaltung. Die Motivation ist, dass der Compiler – soweit möglich – alle Parameter in Registern übergibt und den Stack nur verwendet, wenn sich dies nicht vermeiden lässt.

Aufgabe 4.7 Stack

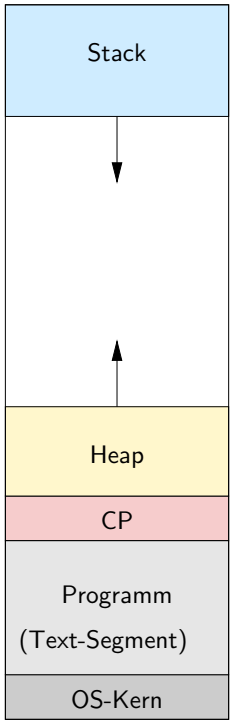
Machen Sie sich aus den Vorlesungsunterlagen die Funktion eines Stacks klar.

(a) Was könnten in diesem Zusammenhang die beiden folgenden Begriffe bedeuten, wenn es um Informationen (z.B. Registerinhalte) geht, die noch benötigt, bzw. überschrieben werden:

- *„caller save“*
- *„callee save“*

(b) Mit welchen Befehlen kann der D·CORE-Stackpointer auf den in Abbildung 3 verwendeten Wert von `0xfffe` initialisiert werden?

(c) Worin liegt unter Berücksichtigung von Abbildung 2 bzw. 3 der konzeptionelle Unterschied, wenn der Stackpointer auf `0xfffe` oder aber auf `0x0` initialisiert wird?



0xffff

```
long fak( n ) {       // Programmtext: Programm
  long n;             // lokale Variable: Stack

  if (n == 1) return 1;
  else        return n*fak(n-1);
}


static int i, j, k;    // globale Variable: CP

int main( int argc; char ** argv ) {
   int* array = malloc( 10000L );// malloc: Heap

  for( i = 0; i < 100; i++ ) {
   printf( "Fakult t(%ld)= %ld", i, fak(i));
  }
}
```

0x0000

Abb. 2: Speicherbereiche im Hauptspeicher: Textsegment, Konstantenpool, Heap, Stack

Aufgabe 4.8 push()

In den meisten Situationen müssen nicht alle, sondern nur einige Register auf den Stack gesichert werden. Notieren Sie als Beispiel die **Assemblerbefehle**, um den Inhalt der Register `R4`, `R5`, `R10` auf den Stack zu sichern. Per Konvention soll Register `R0` als Stackpointer verwendet werden und dieser ist, wie bereits in Aufgabe 4.7 (b) gefordert, auf den Wert `0xfffe` initialisiert.

Aufgabe 4.9 pop()

Notieren Sie die notwendigen **Assemblerbefehle**, um den Inhalt der Register `R4`, `R5`, `R10` vom Stack wiederherzustellen.

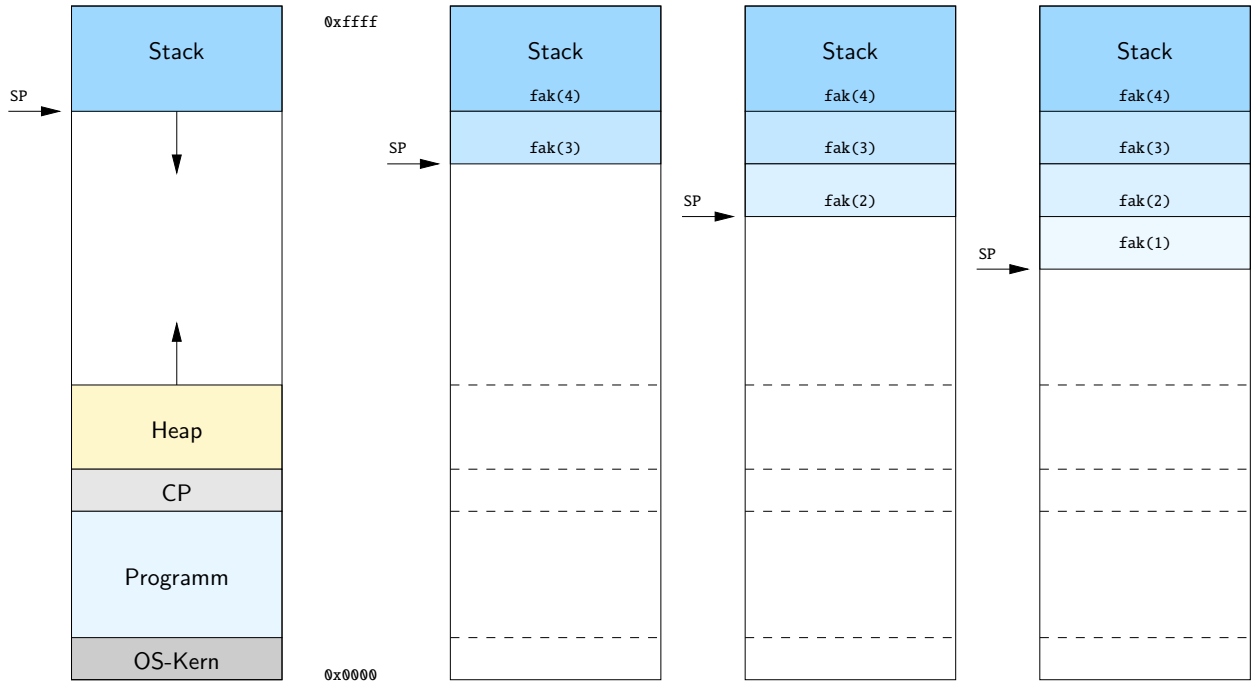


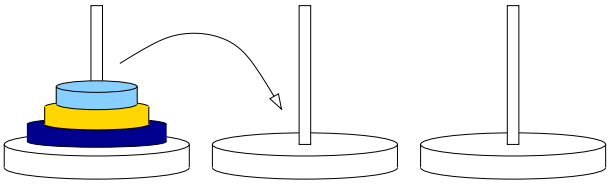
Abb. 3: Rekursiver Aufruf der Fakultätsfunktion.

Stackunterstützung des D·CORE-Assemblers

Viele gute Assembler stellen entsprechende Macros zur Verfügung, wobei die betroffenen Register als Argumente übergeben werden. Das gilt auch für den von uns verwendeten Assembler. Hier heißen die betreffenden Macros `.push` und `.pop`. Als Stackpointer wird als Default das Register `R0` angenommen, das auf die zuletzt beschriebene Speicherstelle zeigt. Falls Sie ein anderes Register als Stackpointer verwenden möchten (z.B. das `R14`), können sie dies dem Assembler mit `.stack R14` mitteilen. Vergessen Sie bitte nicht, ihren Stackpointer auf einen definierten Wert (z.B. 0) zu initialisieren.

Aufgabe 4.10 Rekursive Unterprogramme – Türme von Hanoi

Das Problem der Türme von Hanoi ist eine der bekanntesten Aufgaben mit einer einfachen rekursiven Lösung.



Es geht dabei darum, die Scheiben von Stab 1 auf Stab 3 zu übertragen, wobei jede Scheibe immer auf einem der drei Stäbe liegt und immer eine kleinere Scheibe auf einer größeren liegt. Für drei Scheiben hat man z.B. die sieben Verschiebungen:

$$1 \longrightarrow 3, \quad 1 \longrightarrow 2, \quad 3 \longrightarrow 2, \quad 1 \longrightarrow 3, \quad 2 \longrightarrow 1, \quad 2 \longrightarrow 3, \quad 1 \longrightarrow 3$$

Das folgende Programm zur Lösung des Problems stammt aus [Tanenbaum]:

```
1  #include <stdio.h>
2
3  /*
4    Übertrage n Scheiben von Stab i auf Stab j. (1 <= i, j <= 3).
5  */
6
7    void towers(int n, int i, int j) {
8      if (n == 1) {
9        printf("Übertrage Scheibe von %d nach %d\n", i, j );
10     }
11     else {
12       int k = 6 - i - j;
13       towers(n-1, i, k);
14       towers(  1, i, j);
15       towers(n-1, k, j);
16     }
17   }
18
19 void main() {
20   towers(3, 1, 3);
21 }
```

Realisieren Sie das Programm in D·CORE Assembler und testen Sie es zuerst mit dem angegeben Aufruf `towers(3, 1, 3)`, der zu insgesamt sieben Ausgaben (s.o.) auf dem Terminal führen sollte.

Hinweise zur praktischen Realisierung
Der Assembler kennt drei Pseudobefehle, um etwas auf dem Display des Emulators ausgeben zu können und zwar:

`.prdez` ⟨***reg***⟩ gibt den Inhalt des Registers ⟨*reg*⟩ als Dezimalzahl aus

`.prnewline` gibt einen Zeilenumbruch aus

`.prstr` ⟨***reg***⟩ gibt den String aus, dessen Startadresse im Register ⟨*reg*⟩ steht

Es sei betont, dass es sich hier um eine besondere Form der Pseudoanweisungen handelt, die anders als z.B. .push und .pop durch den Assembler **nicht** auf gültige Befehle bzw. Befehlsfolgen des D·CORE-Prozessors abgebildet werden.
(Siehe hierzu: Aufgabe 4.6, Hinweise zur Pseudoanweisung)

Aufgabe 4.11 Laufzeit

Für größere Parameter wachsen die Laufzeit und der auf dem Stack benötigte Platz schnell an. Erweitern Sie ihre Funktion so, dass die Anzahl der Aufrufe mitgezählt wird. Das Hauptprogramm sollte dann diese Zahl zusammen mit einem String (z.B. *Zahl der Aufrufe*=) auf dem Display ausgeben.

(a) Wieviele Aufrufe ergeben sich für `towers(8,1,3)`?

(b) Was folgt daraus für die Komplexität des Algorithmus?

Aufgabe bearbeitet	abzeichnen lassen