

ARHITECTURA CALCULATOARELOR 2003/2004

CURSUL 13

5.1.4 Microinstrucțiuni

Pentru a controla calea de date din figura 5.1 avem nevoie de 29 de semnale. Acestea pot fi împărțite în 5 grupe funcționale descrise mai jos:

- 9 semnale pentru controlul scrierii datelor de pe magistrala C în regiștri,
- 9 semnale pentru controlul activării regiștrilor pe magistrala B pentru intrarea ALU,
- 8 semnale pentru controlul funcțiilor ALU și ale shifter-ului,
- 2 semnale (nefigurate) pentru indicarea citiri/scrierii memoriei prin MAR/MDR,
- 1 semnal (nefigurat) pentru indicarea extragerii din memorie prin PC/MBR

Valorile acestor 29 de semnale de control specifică operațiile pentru un ciclu al căii de date. Un ciclu constă în direcționarea valorilor din regiștri pe magistrala B, propagarea semnalelor prin ALU și shifter, conducerea lor pe magistrala C și în final scrierea rezultatelor în regiștrul sau regiștrii corespunzători. Dacă un semnal de citire a memoriei este activat, operația cu memoria începe la sfârșitul ciclului căii de date, după ce MAR a fost încărcat. Datele memoriei sunt disponibile la fiecare sfârșit de ciclu următor în MBR sau MDR și pot fi folosite în ciclul de după acesta. Cu alte cuvinte, o citire a memoriei inițiată pe oricare port la sfârșitul ciclului K trimite date care nu pot fi folosite în ciclul K+1 ci numai în ciclul K+2 sau mai târziu.

Presupusul comportament care contrazice varianta intuitivă este explicat de figura diagramei de timp anterioare. Semnalele de control al memoriei nu sunt generate în ciclul 1 al ceasului decât după ce MAR și PC sunt încărcate la frontul crescător al ceasului, către sfârșitul ciclului 1 de ceas. Vom presupune că memoria pune rezultatele pe magistrala memoriei în timpul unui ciclu deci MBR și/sau MDR pot fi încărcate la următorul front crescător al ceasului, împreună cu alți regiștri. Cu alte cuvinte, încărcăm MAR la sfârșitul unui ciclu al căii de date și inițializăm memoria la scurt timp după aceea. Ca urmare, nu ne putem aștepta ca rezultatele unei operațiuni de citire să fie în MDR la începutul următorului ciclu în special dacă lărgimea pulsului ceasului este scurtă. Nu este destul timp dacă memoria mai are nevoie de un ciclu de ceas. Un ciclu al căii de date trebuie să apară între începerea citirii memoriei și folosirea rezultatului. Bineînțeles alte operațiuni pot avea loc în timpul ciclului, dar nu cele care au nevoie de cuvântul de memorie.

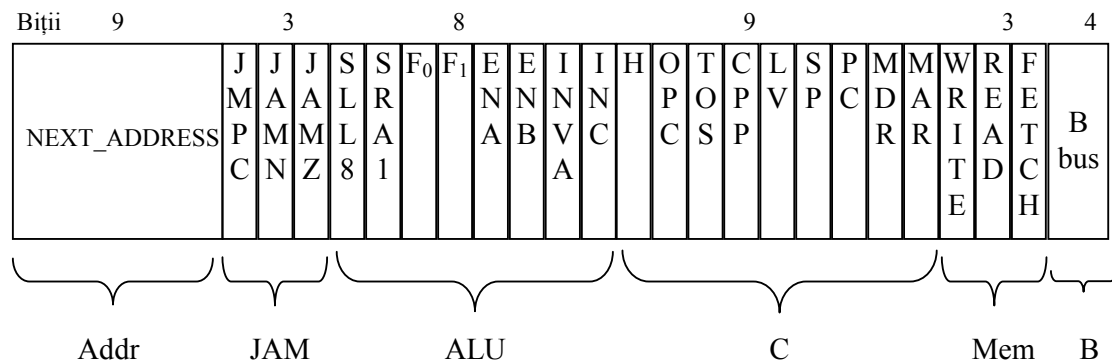
Presupunerea că memoria face un ciclu pentru a opera, este echivalent cu a presupune că rata de operare a cacheului este 100%. Această presupunere nu este adevărată niciodată, dar complexitatea introdusă de un ciclu de memorie cu lungime variabilă este mai mult decât avem nevoie aici. Din vreme ce MBR și MDR sunt încărcate la frontul crescător al ceasului, odată cu toți ceilalți regiștri, aceștia pot fi citiți în timpul ciclurilor când este făcută o nouă citire a memoriei. Acestea returnează vechile valori din moment ce citirea nu a avut încă timp să le rescrie. Nu există nici o ambiguitate aici: până când noile valori sunt încărcate în MBR și MDR la frontul crescător al ceasului, valorile anterioare sunt încă utilizabile. Notați că este posibil să se execute citiri “back to back” în două cicluri consecutive din moment ce o citire are nevoie de doar un ciclu. De asemenea, ambele memorii pot opera în același timp. Totuși încercând să citim și să scriem același octet simultan vom avea rezultate nedefinite.

Dacă se dorește să se scrie ieșirea de pe magistrală C în mai mult de un registru, nu este de dorit să activăm mai mult de un registru la un moment dat pe magistrala B (practic se ard niște circuite dacă se încearcă). Cu o mică creștere în circuite, putem reduce numărul de biți necesari pentru a selecta între posibilele surse pentru conducerea magistralei B. Nu sunt decât 9 regiștrii de intrare posibili care pot conduce magistrala B. De aceea putem coda informația din magistrala B pe 4 biți și putem folosi un decodor pentru a genera cele 16 semnale de control dintre care de 7 nu avem nevoie. Într-un proiect comercial inginerii vor simți nevoia copleșitoare de a scapa de unul dintre regiștri pentru ca 3 biți să facă treaba. Ca profesori ne permitem luxul de a fi în stare să pierdem un bit pentru a prezenta un proiect simplu și curat.

În acest punct putem controla calea de date cu $9+4+8+2+1=24$ de semnale, deci 24 de biți. Oricum acești 24 de biți controlează doar calea de date pentru un ciclu. A doua parte a controlului are sarcina să determine ce va trebui făcut în ciclul următor. Pentru a include aceasta în proiectul controlerului, se va crea un format pentru descrierea operațiilor care vor fi efectuate folosind cei 24 de biți de control plus două câmpuri adiționale. "NEXT_ADDRESS" și "JAM". Conținutul fiecăruia dintre aceste câmpuri îl vom discuta pe scurt.

Figura 5.5 arată un posibil format, împărțit în 6 grupe și conținând următoarele 36 de semnale:

- Addr – conține adresa unei potențiale următoare microinstrucțiuni,
- JAM – determină cum va fi selectată următoarea microinstrucțiune,
- ALU – funcțiile ALU și shifter.
- C – selectează care regiștri vor fi scriși de pe magistrala C,
- Mem – funcțiile memoriei,
- B – selectează sursa magistralei B.



Regiștrii magistralei B

- | | |
|--------|-----------|
| 0=MDR | 5=LV |
| 1=PC | 6=CPP |
| 2=MBR | 7=TOS |
| 3=MBRU | 8=OPC |
| 4=SP | 9-15 none |

Figura 5.5 Formatul microinstrucțiunii

Ordonarea grupurilor este în principiu arbitrară deși defapt s-a ales foarte atent pentru a minimaliza întretăierea între linii într-o figură următoare. Întretăierea liniilor în diagrame schematice, ca în figura în discuție, corespunde deseori cu întretăierea traseelor în cipuri, care cauzează probleme în proiectele bidimensionale și e mai bine să fie evitate.

5.1.5 Controlul microinstrucțiunii: MIC-1

Până acum am descris cum este controlată calea de date, dar încă nu am descris cum se decide care dintre semnalele de control trebuie activate în fiecare ciclu. Aceasta este determinată de un secvențiator (sequencer) care este responsabil de parcurgerea secvențelor de operații necesare pentru executarea unei singure instrucțiuni ISA.

Secvențiatorul trebuie să producă două tipuri de informații pe fiecare ciclu:

- Starea fiecărui semnal de control în sistem.
- Adresa microinstrucțiunii care urmează să fie executată.

5.1.5.1 Funcția de control a secvențiatorului

Figura 5.6 este o diagramă bloc detaliată a unei microarhitecturi complete a mașinii prezentate ca exemplu, pe care o vom numi Mic-1. Când veți înțelege pe deplin fiecare căsuță și fiecare linie în această figură, veți fi pe cale să înțelegeți nivelul microarhitecturii. Diagrama bloc are două părți: calea de date, în stânga, pe care deja am discutat-o în detaliu, și secțiunea de control, în dreapta la care ne vom concentra acum.

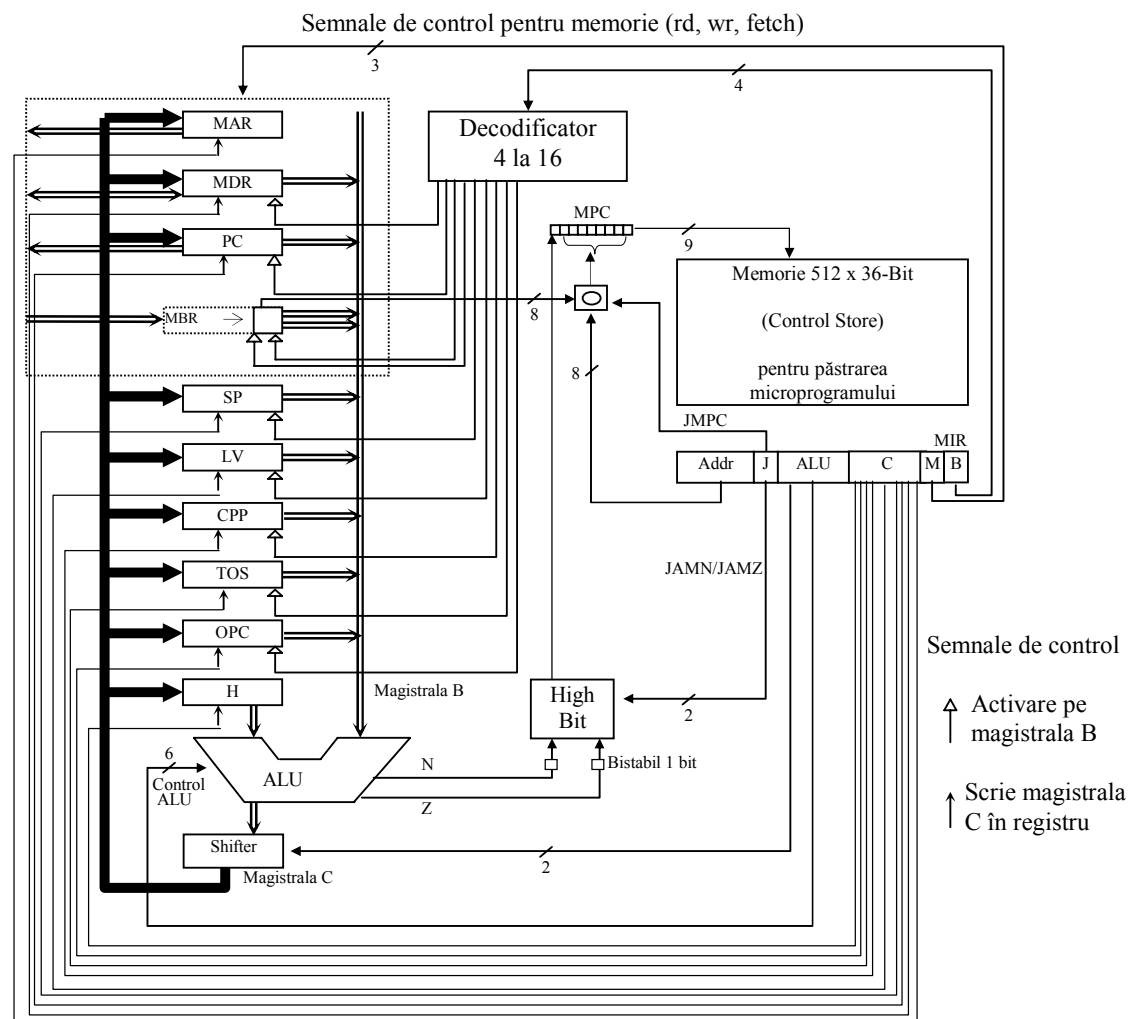


Figura 5.6 Structura microarhitecturii

Cel mai mare și cel mai important lucru din partea de control a mașinii este memoria numită “control store”. Este convenabil să o privim ca o memorie care conține microprogramul complet, deși uneori este implementată ca un set de porți logice. În general, ne vom referi la aceasta cu termenul “control store”, pentru a evita confuzia cu memoria principală, accesată prin MBR și MDR. Oricum funcțional control store este o memorie care pur și simplu conține microinstrucțiuni înloc de instrucțiuni ISA. Pentru mașina luată ca exemplu, conține 512 cuvinte, fiecare cuvânt constând dintr-o microinstrucțiune de 36 de biți ilustrată în figura 5.5.

Într-un mod important, “control store” este diferit de memoria principală: instrucțiunile din memoria principală au proprietatea de a fi executate în ordinea adreselor (cu excepția salturilor), iar microinstrucțiunile **nu**. Incrementarea contorului programului exprimă faptul că instrucțiunea implicită care va fi executată după aceea curentă este instrucțiunea care urmează instrucțiunii curente în memorie. Microprogramele au nevoie de mai multă flexibilitate (pentru că secvențele de microinstrucțiuni au tendința de a fi scurte), deci de obicei nu au această proprietate. În schimb fiecare microinstrucțiune își specifică explicit succesorul. Din vreme ce “control store” este funcțional o memorie (read only), are nevoie de propriul registru de adrese de memorie și de propriul registru de date de memorie. Nu are nevoie de semnale de citire și de scriere, pentru că este continuu citită. Vom numi registrul de adrese de memorie al “control store” MPC (microprogram counter). Acest nume este ironic din moment ce locațiile din el sunt explicit neordonate, deci conceptul de counting nu este util (dar cine suntem noi să punem sub semnul întrebării tradiția?). Registrul de date de memorie se numește MIR, (microinstruction register). Funcția sa este să păstreze microinstrucțiunea curentă ai căror biți conduc (drive) semnalele de control care operează calea de date. Registrul MIR din figura 5.6 conține aceleași 6 grupuri ca în figura 5.5. Grupurile Addr și J (de la JAM) controlează selecția următoarei microinstrucțiuni. Grupul ALU conține 8 biți care selectează funcția ALU și conduc shifterul. Biții C determină regiștrii individuali să încarce ieșirile ALU din magistrala C. Biții M controlează operațiile de memorie.

Ultimii 4 biți conduc decodorul care determină ceea ce merge pe magistrala B. În acest caz s-a ales să se folosească un decodificator standard 4 la 16, deși doar 9 posibilități sunt necesare. Într-un proiect mai detaliat poate fi folosit un decodificator 4 la 9. Varianta comercială este să se folosească un circuit standard luat din biblioteca de circuite pentru a nu construi unul nou. Folosind circuitul standard este simplu și evităm să introducem erori. Folosind propriul cip, cu arie mai mica de utilizare, ia mai mult timp pentru a fi proiectat și s-ar putea să-l greșim.

Modul de funcționare pentru figura 5.6 este prezentat mai jos. La începutul fiecărui ciclu de ceas (frontul descrescător al ceasului din figura 5.3), MIR este încărcat din cuvântul din “control store” indicat de MPC. Timpul de încărcare MIR este indicat în figura 5.3 de Dw. Dacă judecăm în termeni de subcicluri, MIR este încărcat în timpul primului subciclu. Odată ce microinstrucțiunea este încărcată în MIR, variate semnalele sunt propagate pe calea de date. Un registru este scos pe magistrala B, ALU știe ce operație să execute, și va fi multă treabă de făcut acolo. Acesta este al doilea subciclu. După un interval Dw+ Dx de la începutul ciclului intrările ALU sunt stabile. Încă cu Dy mai târziu totul a fost aranjat și ALU, N, Z și ieșirile shifter-ului sunt stabile. Valorile N și Z sunt atunci salvate într-o pereche de bistabili pe un bit. Acești biți, ca și restul de regiștri care sunt încărcăți de pe magistrala C și din memorie, sunt salvați pe frontul crescător al ceasului aproape de sfârșitul ciclului căii de date. Ieșirea ALU este transmisă în shifter. Activitatea ALU și a shifterului au loc în timpul subciclului 3. După un interval adițional Dz ieșirea shifterului a ajuns în regiștri prin magistrala C. Acum regiștrii pot fi încărcăți spre sfârșitul ciclului (la frontul crescător al pulsului ceasului).

Subciclul 4 constă în încărcarea regiștrilor și a bistabilelor N și Z. Se termină puțin după frontul crescător al ceasului, când toate rezultatele au fost salvate și rezultatele anterioarei operații de memorie sunt disponibile și MPC a fost încărcat. Acest proces continuă și continuă până cineva se plictisește și oprește mașina.

5.1.5.2 Determinarea adresei microinstrucțiunii care urmează să fie executată

În paralel cu conducerea căii de date, microprogramul trebuie să determine care instrucțiune trebuie executată în continuare pentru că nu sunt în ordinea în care apar în memoria de comenzi. Calcularea adresei următoarei microinstrucțiuni începe după ce MIR a fost încărcat și este stabil. Câmpul de 9 biți al adresei următoare NEXT_ADDRESS este copiat în MPC. În timp ce copierea are loc JAM este inspectat. Dacă are valoarea 000 nu se întâmplă nimic altceva; când copierea adresei următoare NEXT_ADDRESS este completată, MPC va indica următoarea instrucțiune.

Dacă unul sau mai mulți dintre biții JAM sunt 1 atunci este nevoie de mai multă muncă. Dacă JAMN este setat, bistabilul pe 1 bit N este făcut OR în bitul de rang superior, High Bit, al MPC. Asemănător dacă JAMZ este setat, atunci este făcut OR bistabilul Z pe 1 bit, în High Bit. Dacă amândouă sunt setate atunci ambele sunt făcute OR acolo. Motivul pentru care N și Z sunt necesare este că după trecerea frontului crescător al ceasului (când ceasul este High), magistrala B nu mai este condusă și ieșirile ALU nu mai pot fi considerate corecte. Salvând indicatoarele (flags) ALU în N și Z, se fac disponibile valorile corecte și stabile pentru calculul lui MPC, indiferent de ce se întâmplă în ALU.

În figura 5.6, logica care efectuează aceste calcule este etichetată "HIGH BIT". Funcția booleană realizată este: $F = (JAMZ \text{ AND } Z) \text{ OR } (JAMN \text{ AND } N) \text{ OR } \text{NEXT_ADDRESS} [8]$. În toate cazurile, MPC poate lua numai una dintre cele două valori posibile.

- Valoarea NEXT_ADDRESS.
- Valoarea NEXT_ADDRESS cu HIGH BIT făcut OR cu 1.

Nu există altă posibilitate. Dacă cu bitul high al NEXT_ADDRESS a fost deja făcut 1 atunci folosirea lui JAMN sau JAMZ nu are nici un sens.

Când toți biții lui JAM sunt 0, adresa următoarei microinstrucțiuni ce va fi executată este pur și simplu numărul pe 9 biți din câmpul adresei următoare NEXT_ADDRESS. Când ori JAMN ori JAMZ sunt 1, sunt doi potențiali succesori: NEXT_ADDRESS și NEXT_ADDRESS făcut OR cu 0x100 (presupunând că NEXT_ADDRESS ≤ 0xFF). (0x arată că următorul număr este în sistem hexazecimal). Acest punct este ilustrat în figura 5.7. Microinstrucțiunea curentă la locația 0x75 are NEXT_ADDRESS = 0x92 și JAMZ setat la 1. În consecință următoarea adresă a microinstrucțiunii depinde de bitul Z memorat la operația ALU anterioară. Dacă bitul Z este 0, următoare microinstrucțiune vine din 0x92. Dacă bitul Z este 1, următoare microinstrucțiune vine din 0x192.

Al treilea bit în câmpul JAM este JMPC. Dacă el este setat, cei 8 biți MBR sunt făcuți OR cu cei 8 biți de pondere inferioară ai NEXT_ADDRESS din microinstrucțiunea curentă. Rezultatul este trimis MPC-ului. Căsuța cu eticheta "O" din figura structurii generale face OR din MBR cu NEXT_ADDRESS dacă JMPC este 1, dar decât trece NEXT_ADDRESS spre MPC dacă JMPC este 0. Când JMPC este 1, cei 8 biți de pondere inferioară din NEXT_ADDRESS sunt normal zero. Biții de pondere superioară pot fi 1 sau 0, deci valoarea NEXT_ADDRESS folosită cu JMPC este în mod normal 0x000 sau 0x100.

Abilitatea de a face OR din MBR împreună cu NEXT_ADDRESS și de a stoca rezultatul în MPC permite o implementare eficientă a salturilor ramificate multiplu. De notat că pot fi specificate oricare dintre cele 256 de adrese, determinate numai de biții prezenți în MBR. Într-o folosire tipică MBR conține codul de operație astfel că folosirea JMPM va determina o selecție unică pentru următoarea microinstrucție pentru orice posibil cod de operație. Această metodă este folosită pentru ramificarea rapidă direct la funcția care corespunde codului operației tocmai extras. Înțelegerea secvențierii mașinii este foarte importantă ceea ce face utilă repetarea ei. O vom face făcând referire la subciclii, pentru ca aceasta să fie ușor de vizualizat, dar singurul ceas real al evenimentelor sunt fronturile descrescătoare care încep ciclul și fronturile crescătoare care încarcă registrele și bistabilele N și Z.

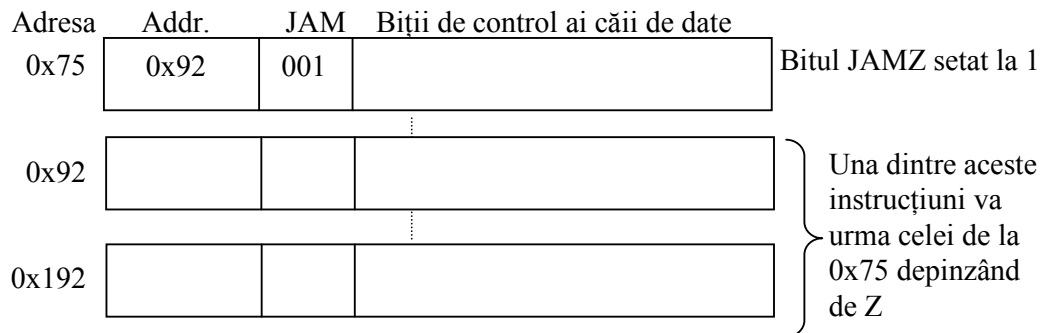


Figura 5.7

În timpul subciclului 1, inițiat de frontul căzător al ceasului, MIR este încărcat din adresa curentă conținută în MPC. În timpul subciclului 2, semnalele primite de la MIR sunt emise și magistrala B este încărcată din registrul selectat. În timpul subciclului 3, ALU și shifterul operează și se produce un rezultat stabil. În timpul subciclului 4, magistrala C, magistralele de memorie, și valorile lui ALU devin stabile. La frontul crescător, registrele sunt încărcate din magistrala C, bistabilele N și Z sunt încărcate, iar MBR și MDR primesc rezultatele de la operația pornită la sfârșitul ciclului de date anterior (dacă a existat). Imediat ce MBR devine disponibil, MPC este încărcat pentru a pregăti următoarea microinstrucție. Astfel MPC își primește valoarea undeva în mijlocul intervalului în care ceasul e high, dar după MBR/MDR sunt gata. Poate fi folosită declanșarea pe nivel (înloc de declanșarea pe front) sau declanșarea la un interval de timp fix față de frontul crescător al ceasului. Tot ceea ce contează este ca MPC să nu fie încărcat până ce registrele de care depinde (MBR, N și Z) nu sunt gata. Imediat ce ceasul cade, MPC poate adresa control store și un nou ciclu poate începe. Remarcați că fiecare ciclu se autoconține. Acesta specifică ce se duce la magistrala B, ce au de făcut ALU și shifterul, unde se stochează magistrala C și care poate să fie următoarea valoare a MPC.

O remarcă utilă despre structura prezentată. Se tratează MPC ca un registru propriu-zis cu o capacitate de memorare de 9 biți care este încărcat în timp ce ceasul este high. În realitate nu este nevoie de a avea un registru. Toate intrările lui pot fi incluse direct în control store. Cât timp aceste intrări sunt prezente în control store la fronturile descrescătoare ale ceasului, când MIR este selectat și citește, înseamnă că este suficient. Nu este nevoie să se stocheze aceste valori în MPC. Pentru acest motiv, MPC poate fi foarte bine implementat ca un registru virtual, care este doar un loc de plasare a semnalelor, mai mult decât un dispozitiv electronic, decât un registru real. A face MPC un registru virtual simplifică secvențierea: acum evenimentele se petrec numai la fronturile descrescătoare și crescătoare ale ceasului nicăieri în altă parte. Dar dacă aceasta este mai ușor de imaginat atunci puteți imagina MPC ca un registru real, și aceasta fiind un punct de vedere valid.

5.2 Exemplu ISA: IJVM

Să continuăm exemplele noastre prin introducerea nivelului ISA al mașinii care trebuie interpretat de microprogramul care rulează pe microarhitectura din paragraful precedent. Pentru ușurință, ne vom referi uneori la Instruction Set Architecture ca la **macroarhitectură**, în contrast cu microarhitectura.

5.2.1 Stive

Virtual toate limbajele de programare suportă conceptul de procedură care are variabile locale. Aceste variabile pot fi accesate din interiorul procedurii, dar încetează să poată fi accesibile odată ce procedura e terminată. Întrebarea ce trebuie pusă este: “Unde trebuie stocate în memorie aceste variabile?”

Cea mai simplă soluție, să se dea oricărei variabile o adresă absolută de memorie, nu este bună. Problema este că procedura se poate autoapela singură. Pentru moment este suficient să spunem că dacă o procedură este apelată de două ori, este imposibil de memorat variabilele ei în locații de memorie absolute pentru că a doua apelare va interfera cu prima. Alternativ, este folosită o strategie diferită. O zonă a memoriei, numită stivă, este rezervată variabilelor, iar variabilele individuale nu conțin adrese absolute. Un registru, să zicem LV, este setat să indice baza stivei pentru variabilele locale pentru procedura curentă.

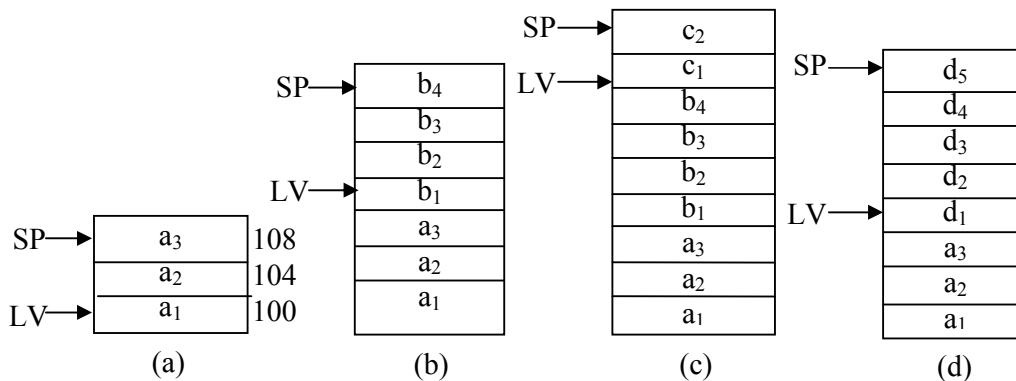


Figura 5.8 Folosirea stivei pentru memorarea variabilelor locale. (a) Când A este activ. (b) După ce A apelează B. (c) După ce B apelează C. (d) După ce C și B s-au terminat și A apelează D.

În figura 5.8 (a) de mai sus, procedura A care are ca variabile locale a_1 , a_2 și a_3 , a fost apelată deci stocarea variabilele locale a fost făcută în zona de memorie rezervată la adresa indicată de LV. Alt registru, SP, indică cea mai mare adresă de cuvânt folosită de variabilele locale ale procedurii A. Dacă LV este 100 și cuvintele au 4 octeți, atunci SP va fi 108. Variabilele sunt referite dându-se adresa lor relativă la baza indicată de LV. Structura de date dintre LV și SP, inclusiv ambele cuvinte indicate de cei doi regiștrii, se numește **structura variabilelor locale** a procedurii A.

Acum să considerăm ce s-ar întâmpla dacă A ar apela altă procedură, și anume B. Unde ar trebui B să memoreze 4 variabile locale (b_1 , b_2 , b_3 , b_4)? Răspuns: în stivă, în continuare după A, cum arată figura 5.8 (b). Notați că LV a fost ajustat de către procedura apelată astfel încât

să indice către variabilele locale ale procedurii B în locul acelor ale procedurii A. Variabilele locale ale procedurii B pot fi referite dându-li-se distanța față de LV. Similar, dacă B apelează C, LV și SP sunt ajustate din nou pentru a indica spațiul de memorie pentru cele două variabile locale ale lui C, cum este ilustrat în figura 5.8 (c).

Când C se termina (returns) B devine din nou activ, și stiva este ajustată la loc ca în figura 5.8 (b) deci acum LV indică din nou adresa variabilelor locale ale procedurii B. După ce se termină și B, LV indică din nou adresa variabilelor locale ale procedurii A, ca în figura 5.8 (a). În orice caz LV indică adresa stivei pentru procedurile active la momentul curent și SP indică vârful structurii stivei. Acum presupunem că A apelează D care are cinci variabile locale. Avem situația din figura 5.8 (d) în care variabilele locale ale procedurii D folosesc aceeași memorie pe care o folosește și procedura B, ca și procedura C. Cu această organizare a memoriei, memoria este alocată numai pentru procedurile care sunt active curent. Când procedura se termină, memoria folosită de variabilele locale este eliberată.

Stivele au și o altă folosire, în plus față de păstrarea variabilelor locale. Pot fi folosite pentru păstrarea operanzilor în timpul calculării expresiilor aritmetice. Când o stivă este folosită în acest fel se spune că este o stivă de operanzi.

Presupunând, de exemplu, că înainte de a se apela B, A are de calculat: $a_1 = a_2 + a_3$; O cale de a face această sumă este de a pune (push) a_2 în stivă ca în figura 5.9 (a). Aici SP a fost incrementat cu numărul de octeți din cuvânt, să zicem 4 și primul operand este memorat la adresa indicată acum de SP. Apoi a_3 este plasat (push) în stivă ca în figura 5.9 (b). Numele variabilelor și ale procedurilor sunt alese de utilizator, codul operației și numele registrilor sunt prestabilite. Calculul poate fi executat acum lansând o instrucțiune care extrage (pop) două cuvinte din stivă, le adună și pune (push) rezultatul în stivă ca în figura 5.9 (c). În final, cuvântul din vârful stivei poate fi extras (pop) și rememorat la loc în variabila locală a_1 ca în figura 5.9 (d).

Stivele pentru variabile locale și acelea pentru operanzi pot fi amestecate între ele. De exemplu când calculăm expresia $x^2 + f(x)$, o parte a expresiei (de exemplu x^2) poate fi stocată în stiva de operanzi când funcția f este apelată. Rezultatul funcției rămâne în stivă, deasupra lui x^2 , deci următoarea instrucțiune poate aduna la această valoare.

În timp ce toate mașinile folosesc stiva pentru memorarea variabilelor locale, nu toate aceste mașini folosesc stivele pentru calcule aritmetice. De fapt cele mai multe dintre ele nu le folosesc, dar JVM și IJVM funcționează astfel ceea ce justifică prezentarea acestor operații aici.

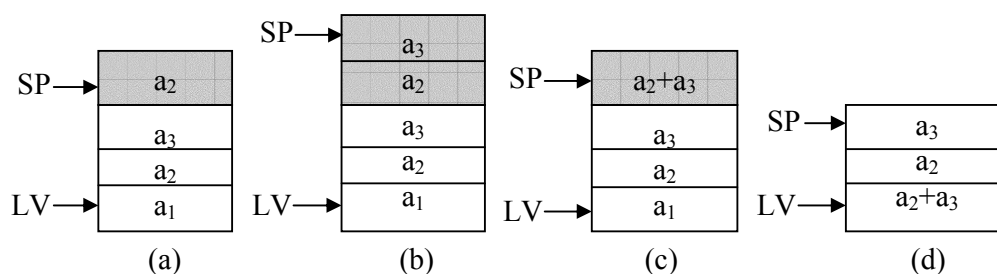


Figura 5.9 Folosirea stivele de operanzi pentru realizarea operațiilor aritmetice.