Nanotehnologia: o soluţie pentru calculatoarele viitorului

Mihai Budiu — mailto:mihaib+@cs.cmu.edu
http://www.cs.cmu.edu/~mihaib/

aprilie 2001

Subiect: o posibilă utilizare a nanotehnologiei pentru construcţia calculatoarelor

Cunoştinţe necesare: cunoştinţe elementare de arhitectura calculatoarelor

Cuvinte cheie: nanotehnologie, hardware reconfigurabil

Cuprins

1 Sfîrşitul erei legii lui Moore

Un adevăr incontestabil este că evoluţia tehnologiilor de calcul a avut o acceleraţie greu de imaginat. Timp de cel puţin 40 de ani, de la inventarea circuitelor integrate, rata de creştere a performanţei a fost exponenţială, dublînd puterea de calcul la fiecare 18 luni. Această evoluţie a fost atît de inexorabilă încît a fost denumită "lege": aceasta este celebra lege a lui Moore.

Termenul poate însă fi înşelător: legea lui Moore este doar o observaţie empirică, şi nu o lege a naturii. Reputaţia ei a fost întărită de faptul că adesea experţii au prezis bariere care vor stăvili avîntul tehnologiei şi vor îi încetini evoluţia. De fiecare dată, însă, realitatea a depăşit predicţiile, şi legea lui Moore a funcţionat timp 40 de ani. Limitele finale sunt însă din ce în ce mai aproape. În mod surprinzător, cele mai ameninţătoare bariere care se profilează sunt de natură economică!

Dacă urmărim costul unei fabrici care produce circuite integrate vom observa că urmează aceeaşi curbă exponenţială pe care o urmează performanţa circuitelor integrate. O astfel de fabrică a ajuns la ora actuală să coste cam 2 miliarde de dolari, şi la aceeaşi rată de creştere va ajunge la 50 de miliarde în 2010. Acesta e un cost pe care foarte puţine companii şi-l pot permite, mai ales că investiţia nu este lipsită de riscuri (vedeţi de exemplu războiul Intel-AMD, cu tot felul de întorsături neaşteptate în poziţia de lider al pieţei).

Chiar dacă putem argumenta că fabricile vor putea fi construite în comun de către consorţii de fabricanţi, există o serie de costuri care nu pot fi distribuite: fiecare circuit integrat este construit folosind o serie de măşti; costul măştilor a suferit şi el o evoluţie exponenţială, făcînd deja prohibitivă fabricarea de chip-uri de către companiile mici.

Creşterea costului de fabricaţie este dată de costul dispozitivelor *mecanice* care sunt folosite în fabricarea integratelor. Pe măsură ce dimensiunea tranzistoarelor scade, cerinţele de precizie pentru alinierea măştilor, şlefuirea suprafeţei plăcii de siliciu, montarea nivelelor de metal ale circuitului final, etc., cresc şi ele; costul unor dispozitive mecanice care fac operaţii atît de precise creşte deci exponenţial cu scăderea dimensiunii.

Anticipînd sfîrşitul acestei ere prodigioase din istoria calculatoarelor, grupul de cercetare din care fac şi eu parte a început să investigheze modele alternative de calcul. În acest articol voi prezenta pe scurt propunerea noastră pentru următoarea generaţie arhitecturală, care substituie în locul circuitelor integrate în tehnologie CMOS (Complementary Metal-Oxide Semiconductor) circuite construite folosind nanotehnologie electronică.

2 Nanotehnologie

Dicţionarul Webster defineşte cuvîntul "nanotehnologie" ca "arta manipulării unor dispozitive minuscule, de dimensiuni moleculare". Guvernul american însă a investit anul trecut jumătate de miliard de dolari în cercetarea din domeniul nanotehnologiilor, în încercarea de a transforma-o dintr-o "artă" într-o "ştiinţă". Ca urmare, activitatea de cercetare din domeniu este prodigioasă; în secţiunea despre alte surse de informaţie ofer cîteva legături care vă pot ghida spre paginile de web ale unora dintre cele mai renumite laboratoare. Progresele rapide, mai ales din chimie, au pus la-ndemîna cercetătorilor o seamă de unelte miniaturale minunate.

În această secţiune voi trece în revistă unele dintre dispozitivele minuscule care se anunţă a fi potrivite pe post de componente ale unui dispozitiv de calcul; nanotehnologia ca domeniu se ocupă de multe alte lucruri, pe care le voi ignora. Chiar şi printre dispozitivele nano-electronice, este imposibil de făcut o enumerare exhaustivă, aşa că mă voi mulţumi doar să ilustrez cu exemple reprezentative.

2.1 Componente

Primul ingredient necesar pentru a construi circuite la scară moleculară este… sîrma. Astfel de sîrme trebuie să fie foarte subţiri, lungi, rezistente mecanic, şi să aibă o conductanţă electrică bună. Din fericire chimiştii au descoperit o serie de molecule care au exact proprietăţile necesare. Figura 1 este ilustraţia unui fragment din cea mai celebră dintre moleculele descoperite, nanotubul de carbon.

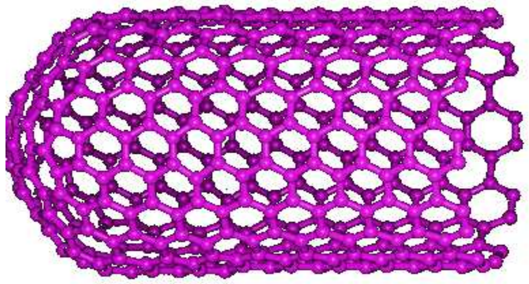

Figura 1: Nanotuburi de carbon: fiecare biluţă este un atom de carbon legat covalent cu vecinii săi. Aceasta este partea din stînga a *unei singure molecule*. Pentru descoperirea acestor molecule Richard Smaley, de la universitatea Rice din Statele Unite, a primit premiul Nobel pentru chimie în 1996. Astfel de molecule au un diametru de 5 nm şi pot avea lungimi de ordinul milimetrilor. Proprietăţile lor electrice şi mecanice sunt excelente, în pofida dimensiunilor lor minuscule.

Al doilea ingredient de care avem nevoie este un comutator, care poate închide şi deschide circuite. Din fericire şi pentru acest dispozitiv există o pletoră de alternative; în figura 2(a) este schema unui astfel de comutator; fiecare biluţă este un atom.

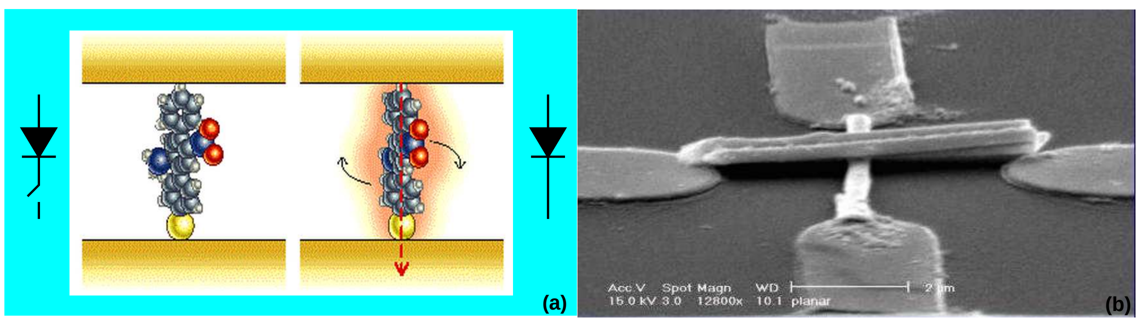

Figura 2: (a) Un comutator molecular. O moleculă polarizată are un nor de sarcină electrică asimetric. În poziţia din stînga norul blochează trecerea curentului electric. Aplicînd un potenţial ridicat cauzăm rotirea moleculei şi reorientarea norului electronic; molecula în poziţia din dreapta conduce curent electric într-o singură direcţie, comportîndu-se ca o diodă. Molecula are "memorie", pentru că rămîne pentru multă vreme în poziţia în care a fost pusă. Un potenţial mare negativ poate muta molecula înapoi în starea neconducătoare. [această figură este adaptată din revista Scientific American] (b) Comutatorul molecular plasat între două nano-sîrme.

În figura 2(b) avem micro-fotografia unui astfel de comutator cuplat cu două nano-sîrme. Cum putem poziţiona atît de precis comutatorul? Prin procedee chimice simple:

1. Fabricăm cele două sîrme separat;
2. Înmuiem una dintre sîrme într-o soluţie care conţine molecule-comutator;
3. Suprapunem cele două sîrme aproximativ în unghi drept.

Astfel, comutatorul de la intersecţia celor sîrme se va cupla de ambele şi va deveni funcţional.

Faptul că putem construi sîrme izolate şi legate prin comutatoare nu este însă suficient pentru a construi circuite complexe. Trebuie sa fim capabili să construim în mod eficient (în paralel) multe astfel de sîrme cuplate cu comutatoare. Din fericire chimiştii au descoperit un fenomen care ne oferă soluţia într-un mod aproape miraculos. Acest fenomen se numeşte *auto-asamblare* (self-assembly). Una din formele sale se manifestă astfel: facem o soluţie cu un anumit tip de molecule. În soluţie muiem un suport. Încălzim soluţia, extragem suportul şi în mod spontan, fără vreun control dinafară, moleculele din soluţie se aşează pe substrat într-o structură aproape regulată, paralele între ele. În felul acesta putem construi simultan zeci sau sute de sîrme paralele, aflate la distanţe foarte mici una de alta.

2.2 Trăsături

Auto-asamblarea este un procedeu foarte ieftin de construcţie a unor structuri minuscule. Cu toate acestea, trebuie să realizăm cîteva dintre limitările sale inerente:

- Folosind auto-asamblare nu putem construi structuri neregulate, aperiodice. Circuitele integrate fabricate prin tehnologie CMOS sunt însă structuri neregulate. Va trebui deci să renunţăm la modelul circuitelor integrate digitale CMOS.

- Cel puţin în viitorul apropiat, folosind auto-asamblare afirmăm că este practic imposibil de construit un dispozitiv cu trei terminale, cum este tranzistorul. La dimensiuni moleculare este imposibil de coordonat simultan poziţia a trei sîrme pentru a le face să se întîlnească aproximativ în acelaşi loc. Folosind manipulări minuţioase sub un microscop electronic putem construi un tranzistor, dar afirmaţia noastră este că nu putem construi în mod automat şi paralel milioane de tranzistori. Tranzistorul este însă ingredientul fundamental al circuitelor integrate digitale CMOS, care nu pot fi concepute în absenţa sa! Cum vom putea atunci construi circuite integrate digitale?

  Pentru a putea răspunde la această întrebare trebuie din nou să renunţăm la metodele tradiţionale. Din fericire răspunsul se poate găsi dacă răsfoim publicaţii vechi, îngropate în praf prin biblioteci: la începutul aniilor '50, înainte de inventarea tranzistorului, pentru o vreme cercetătorii au studiat o metodă alternativă de a construi porţi logice, folosind numai diode şi rezistenţe. Figura 3(a) arată implementarea unei porţi logice "şi" folosind diode şi rezistori, pe cînd figura 3(b) arată cum acest circuit poate fi implementat folosind nanotehnologia. Acest mod de a construi circuite a fost abandonat după apariţia tranzistorului, pentru că consumă prea mult curent şi este mai ineficient. Cu toate acestea, consumul dispozitivelor nanoelectronice este minuscul, putînd face acest model viabil din nou.

- În circuitele integrate digitale tranzistorul are un rol dublu: serveşte la implementarea porţilor logice şi amplifică semnalul electric care se degradează. Am văzut că prima funcţionalitate poate fi înlocuită în nano-circuite folosind diode. Avem însă nevoie şi de amplificatoare la scară moleculară. Grupul de cercetare din care fac parte a propus un astfel de dispozitiv care are simultan trei funcţii:

  - Amplifică semnalul electric restaurîndu-l la valoarea normală;

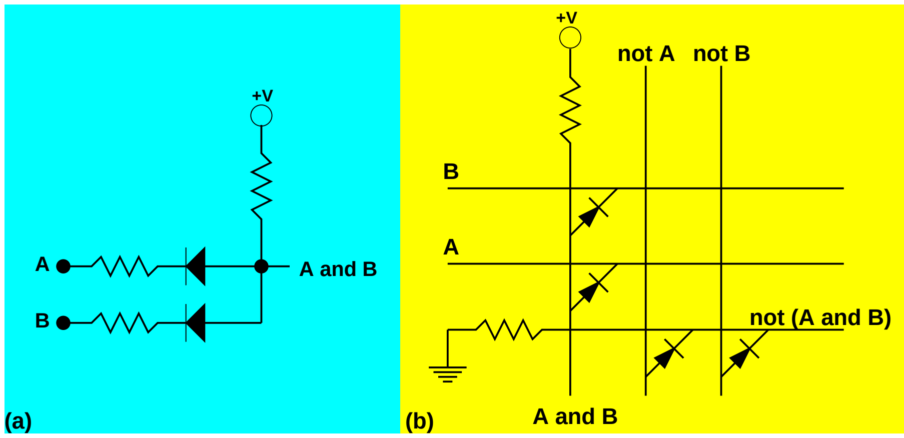

Figura 3: (a) O poartă logică "şi" implementată cu diode şi rezistenţe. Cînd una dintre intrări are tensiune joasă, dioda corespunzătoare conduce, deci tensiunea la ieşire este şi ea joasă. Cînd ambele intrări au un potenţial ridicat, nici una din diode nu conduce, şi potenţialul la ieşire este ridicat. Dacă asociem un potenţial ridicat cu valoarea logica "1", am obţinut astfel funcţionalitatea unei porţi logice "şi". (b) Implementarea porţii logice folosind nano-sîrme şi nano-comutatoare. Observaţi structura regulată şi folosirea exclusivă a sîrmelor care se încrucişează aproximativ la unghi drept.

   - Oferă sincronizare cu un semnal de ceas;
   - Poate fi folosit pe post de element de memorie.

  Acesta este echivalentul molecular al unui "latch" (un registru).

- În fine, cînd avem structuri atît de mici este inevitabil să avem o cantitate ne-neglijabilă de defecţiuni. Mici fluctuaţii termodinamice, un grăunte de praf, vor perturba imediat regularitatea sîrmelor şi comutatoarelor. Circuitele integrate digitale clasice trebuie să fie perfecte; după fabricaţie fiecare este testat extensiv şi cele cu defecţiuni sunt aruncate. Această metodologie va fi inaplicabilă nano-circuitelor, pentru că densitatea de defecţiuni va face existenţa unui circuit perfect practic imposibilă. În secţiunea următoare vom sugera însă o soluţie pentru această ultimă problemă.

3 Hardware reconfigurabil

În PC Report din iulie 1998 am publicat un articol despre o nouă tehnologie de construire a calculatoarelor care se numeşte hardware reconfigurabil. Voi rezuma aici cele mai importante elemente din acel text, dar puteţi accesa şi versiunea sa on-line.

Hardware-ul reconfigurabil are nişte trăsături extrem de dezirabile, cum ar fi posibilitatea de a tolera defecţiuni şi capacitatea de a oferi performanţă computaţională foarte ridicată. Un dezavantaj al hardware-ului reconfigurabil implementat în tehnologie CMOS este că are o densitate scăzută (adică se pot implementa mult mai puţine elemente computaţionale pe aceeaşi suprafaţă decît folosind hardware tradiţional). Sugestia noastră este de a implementa hardware reconfigurabil pe un substrat nanotehnologic. În felul acesta putem obţine beneficiile ambelor domenii şi putem evita dezavantajele lor:

- Datorită dimensiunii foarte reduse a componentelor, nano-hardware-ul compensează densitatea scăzută a dispozitivelor reconfigurabile.
- Datorită toleranţei la defecte, hardware-ul reconfigurabil poate exploata fără probleme nanocircuitele cu defecţiuni.
- În plus, dorim să exploatăm viteza de execuţie excelentă a hardware-ului reconfigurabil.

Ce este hardware-ul reconfigurabil?

Un circuit reconfigurabil (vedeţi figura 4) conţine, ca şi un circuit digital normal, porţi logice şi sîrme. Spre deosebire de circuitele obişnuite însă, fiecare poartă reconfigurabilă nu are o funcţionalitate fixată, ci poate fi re-programată electronic să implementeze orice funcţie logică dorită. În plus, sîrmele din hardware-ul reconfigurabil nu sunt conectate direct între porţi; între sîrme avem *comutatoare* care pot lega şi dezlega sîrmele între ele. Fiecare comutator este controlat tot în mod electronic.

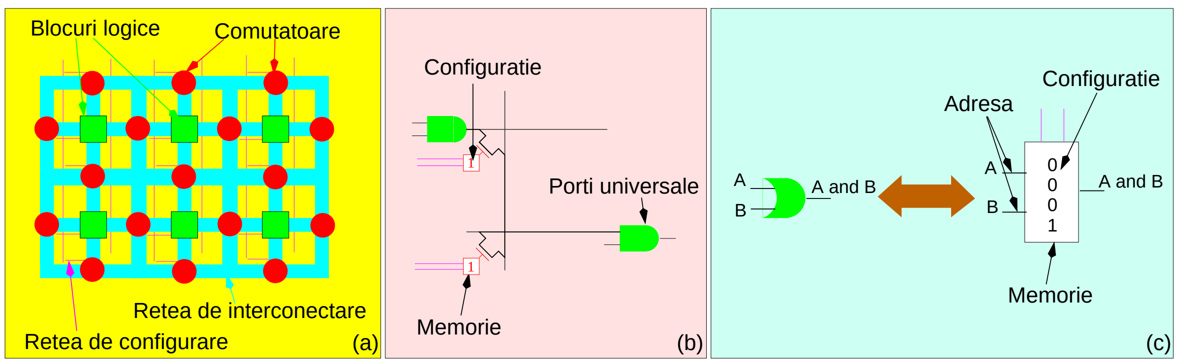

Figura 4: (a) Hardware-ul reconfigurabil constă dintr-o "mare" de porţi logice universale legate cu o reţea de interconectare. O reţea de configurare controlează funcţiunea fiecărei porţi logice şi poziţia fiecărui comutator. (b) Fiecare comutator este controlat de o memorie de 1 bit care indică poziţia sa de închis/deschis. (b) Porţile logice universale sunt implementate folosind memorii mici care funcţionează ca tabele de adevăr: modificînd conţinutul memoriei schimbăm funcţionalitatea porţii logice. În acest exemplu memoria implementează o poartă logica "şi".

Hardware-ul reconfigurabil este deci un caz particular de circuit integrat.

Atunci cînd comparăm un circuit reconfigurabil cu unul normal care are exact aceeaşi funcţiune, cel reconfigurabil este mai lent şi mult mai mare (cam de 10 ori mai mare). Este uşor de înţeles de ce: o poartă logică implementată cu o memorie este mult mai mare decît aceeaşi poartă făcută din tranzistori, iar comutatoarele şi memoriile lor aferente ocupă spaţiu suplimentar. Viteza lui mai scăzută este dată de comutatoare, care încetinesc semnalul electric.

Avantajul hardware-ului reconfigurabil vine din flexibilitatea sa: putem schimba funcţiunea lui pentru fiecare nouă aplicaţie cu care avem de-a face. Microprocesoarele oferă şi ele flexibilitate, dar sunt în general mai ineficiente decît hardware-ul reconfigurabil, din două motive:

- Un procesor *interpretează* programele scrise în cod maşină, pe cînd hardware-ul reconfigurabil le execută direct. Interpretarea introduce ineficienţă.

- Programele executate de un procesor trebuie scrise folosind un set fixat de instrucţiuni (în cod maşină), care nu este neapărat ideal pentru orice aplicaţie. Hardware-ul reconfigurabil poate fi adaptat special fiecărei aplicaţii (de exemplu dacă facem operaţii pe 5 biţi nu e nevoie să folosim mereu cîte 32).

În plus, putem folosi capacitatea de reconfigurare pentru a tolera defecte. Să presupunem că una dintre porţile universale este defectă din fabricaţie. Dacă ştim acest lucru, vom aranja ca nici unul din programele noastre să nu o folosească. Pentru că porţile sunt toate echivalente, putem folosi oricare alta, pe care o putem conecta în locul celei defecte folosind reţeaua configurabilă.

Trebuie să spunem cîteva cuvinte şi despre compilarea pentru hardware reconfigurabil:

- Compilarea se face în două etape: întîi programul este tradus într-un circuit, după care circuitul trebuie "aşezat" în plan, atribuind fiecărei operaţii una sau mai multe poarţi universale şi atribuind sîrmelor interconexiuni. Această ultimă fază se numeşte "plasare şi rutare" (place and route) şi este de obicei foarte costisitoare şi complexă (ore sau zile întregi). Rezultate mai recente ale cercetării (unele obţinute chiar de autorul acestui articol) demonstrează însă că această fază poate fi accelerată substanţial, pînă la viteze comparabile cu compilarea normală de programe.

- În general, problema transformării eficiente a unui program dintr-un limbaj de nivel înalt în hardware este încă nerezolvată satisfăcător. Un motiv constă în faptul că limbajele contemporane au fost proiectate pentru un model de execuţie von Neumann, în care maşina are procesor şi memorie. Cercetarea curentă a autorului acestui articol se orientează în această direcţie.

4 Nano-calculatoare

În această secţiune voi schiţa cum putem pune cap la cap nanotehnologia şi hardware-ul reconfigurabil pentru a construi maşini de calcul. Din informaţiile autorului, grupul de cercetare din care face parte este singurul din lume care atacă problema viitorului calculatoarelor din acest unghi.

Figura 5 arată propunerea noastră de arhitectură a unui sistem de calcul bazat pe nanotehnologii. Structura aceasta este foarte asemănătoare cu cea a circuitelor reconfigurabile disponibile comercial (numite FPGA, Field Programmable Gate Array), pentru care există deja multă experienţă în proiectarea sculelor şi tehnologiilor de compilare, plasare şi rutare.

Astfel de circuite vor fi fabricate printr-o mixtură de tehnologii: fiecare nano-bloc e fabricat folosind nanotehnologii. Lăcaşurile pentru grupuri şi sîrmele lungi sunt fabricate în tehnologie CMOS; fiecare nanobloc este apoi inserat într-un astfel de locaş. Metode prin care se pot configura nano-blocurile au fost puse la punct.

După fabricaţie circuitele vor fi cuplate la un calculator care va testa cîteva din grupuri, pentru a găsi unul perfect funcţional. După aceea calculatorul va configura acest grup cu un program de auto-testare, prin care restul circuitului este testat pentru a găsi alte defecte. Nano-blocurile defecte sunt înregistrate într-o listă de defecte ataşată circuitului. Cînd circuitul este utilizat, compilatorul

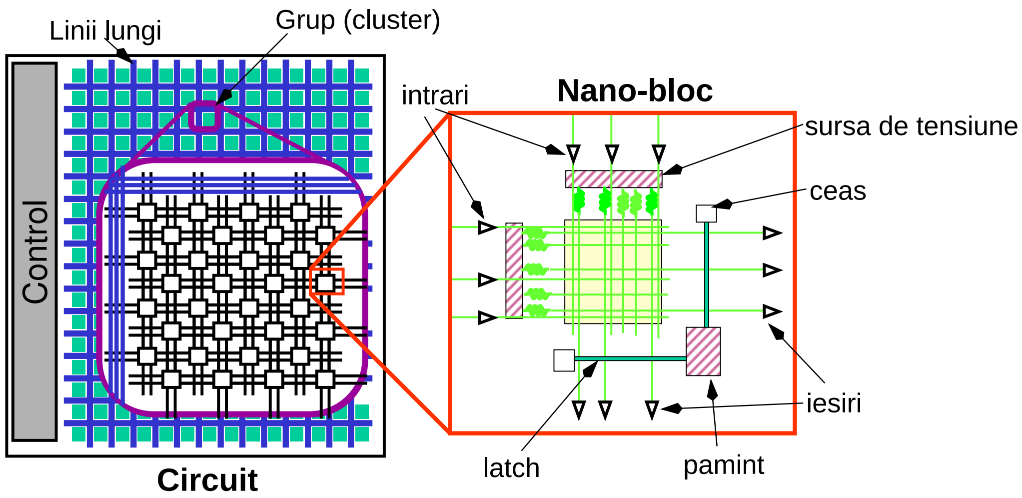

Figura 5: Stînga: Structura unui nano-circuit: un astfel de circuit constă într-o grilă bidimensională de grupuri (clusters), legate prin "sîrme" configurabile de felurite lungimi. Fiecare grup constă dintr-o grilă de nano-blocuri. Dreapta: Fiecare nano-bloc este o poartă logică universală cu trei intrări şi trei ieşiri (fiecare semnal este atît direct cît şi complementat, deci avem cîte şase sîrme). Nano-blocul conţine şi conexiuni la porţiuni de metal construite prin tehnologie CMOS: sursa de putere, pămînt, ceas.

care generează configuraţia va folosi această listă de defecte pentru a genera un circuit care foloseşte numai părţile funcţionale.

5 Concluzii

Domeniul nanotehnologiilor este abia la început, dar promite avansuri tehnologice extraordinare, rata descoperirilor fiind într-o continuă acceleraţie. Se speră ca în circa cinci ani să se poată deja construi nano-circuite digitale funcţionale.

În acest articol am prezentat una dintre propunerile cele mai articulate despre cum ar trebui să arate sistemele de calcul bazate pe nanotehnologie. Am încercat să vă conving că paradigma hardware-ului reconfigurabil este ideală pentru folosirea într-un cadru nano-tehnologic.

Sistemul de calcul al viitorului va fi constituit probabil dintr-un hibrid din siliciu şi nano-circuite. În viitorul ceva mai îndepărtat vom putea probabil folosi şi dispozitive biologice (celor care sunt sceptici de viitorul nanotehnologiei este suficient să le amintim că celulele vii fac calcule sofisticate folosind molecule de ADN). Aşa cum se întîmplă însă adesea, e posibil ca realizările viitorului să întreacă şi cele mai îndrăzneţe preziceri ale noastre.

6 Alte surse de informaţie

- Un articol excelent despre nanotehnologie în limba engleză este disponibil în varianta on-line a revistei Scientific American: “Computing with Molecules”, de Mark A. Reed şi James M. Tour la `http://www.sciam.com/2000/0600issue/0600reed.html`.

- O pagină cu multe legături despre nanotehnologie se află la fostul centru de nanotehnologie al companiei Xerox, acum devenit o companie independentă: `http://www.zyvex.com/nano`.
- Articolul meu despre hardware reconfigurabil din PC Report iulie 1998 este la `http://www.cs.cmu.edu/~mihaib/articles/articles.html#reconf`.
- Un domeniu alternativ este cel al calculatoarelor biologice, bazate pe DNA. O pagină cu cîteva legături despre calculatoare bazate pe DNA: `http://dope.caltech.edu/winfree/DNA.html`.
- Proiectul de calculatoare amorfe de la MIT dezvoltă simulatoare şi limbaje de programare pentru calculatoare autonome foarte simple de dimensiunea unui microb: `http://www.swiss.ai.mit.edu/projects/amorphous/index.html`.
- O copie rătăcită a unui articol din revista Science despre Teramac, cel mai “defect” calculator reconfigurabil care funcţionează perfect, construit de centrul de cercetări de la Hewlett-Packard (revista Science on-line este disponibilă numai contra cost): `http://www.ai.mit.edu/projects/aries/course/notes/teramac.pdf`.
- Un articol foarte interesant despre viitorul calculatoarelor după “epoca siliciului”, scris de unul dintre autorii sistemului Teramac: `http://fips-server.physik.uni-kl.de/fips/teilnehmer/newsticker/history/00-10-31.html`.
- Un articol la care a colaborat şi autorul celui de faţă, care va fi prezentat în vară la conferinţă *International Symposium of Computer Architecture*; acest articol descrie ceva mai amănunţit propunerea noastră arhitecturală precum şi o evaluare preliminară a performanţei unui astfel de model de calcul: `http://www.cs.cmu.edu/~mihaib/research/isca01.ps.gz`.
- Transparentele de la prezentarea mea din cadrul unui seminar de la universitatea Carnegie Mellon, pe 20 aprilie 2001, cu subiectul “Nanotehnologia în slujba calculatoarelor”: `http://www.cs.cmu.edu/~mihaib/research/sss01.ppt` sau `http://www.cs.cmu.edu/~mihaib/research/sss01-slides/slides1.html`.
- Numărul din luna mai al revistei Technology Review publicate de Universitatea MIT este consacrat viitorului calculatoarelor: `http://www.techreview.com/magazine/may00/mag_toc.asp`.