

Veljko MILUTINOVIĆ,* Jakob SALOM**

SINERGIZOVANJE ČETIRI RAZLIČITE RAČUNARSKÉ PARADIGME ZA MAŠINSKO UČENJE I ANALITIKU VELIKIH PODATAKA

Sažetak: Ovaj članak predstavlja i analizira četiri računarske paradigme, koje su prisutne u današnjem svetu IT programiranja — Control Flow, Data Flow, Diffusion Flow i Energy Flow. Porede se njihova glavna svojstva, ukazuje na svrhu svakog od njih i opisuje koje su njihove prednosti i mane. U trećem delu ovog članka autori spekuliraju o mogućoj arhitekturi superkompjuteru na čipu, a u četvrtom predlažu optimalnu raspodelu resursa za određeni skup aplikacija u građevinarstvu.

Ključne reči: *Data Flow, Control Flow, Diffusion Flow, Energy Flow, Maxeler DFE, WSN, BioMolecular computing, QuantumMechanical computing, kompjuterska paradigma*

1. UVOD

Računarska scena danas uključuje četiri različite računarske paradigme i srodne modele programiranja. Neke od paradigmi i modela su u usponu, a druge su na stabilnim osnovama. Ove četiri paradigme su Control Flow (MultiCore — sa više jezgara kao kod Intela i ManyCore — sa puno jezgara kao kod NVidie), Data Flow (fiksni ASIC¹ — bazirani kao kod Google TPU² -a i fleksibilni FPA³ — bazirani kao u početku sa Maxeler DFE⁴, a u poslednje vreme i sa mnogim drugim), Diffusion Flow (npr. sa IoT-om

* Veljko Milutinović, Indiana University, Bloomington, USA, inostrani član Crnogorske akademije nauka i umjetnosti

** Jakob Salom, IPSI Ltd, Belgrade

¹ ASIC — Application-Specific Integrated Circuit

² TPU — Tensor Processing Units

³ FPA — Programabilni hardver

⁴ DFE — Data Flow Engine (mašina)

— internetom stvari i VSN-ima — bežičnim senzorskim mrežama) i Energy Flow — protok energije (kao kod biomolekularnog i kvantnomehaničkog računarstva). Za više detalja pogledajte reference [1...9]

Svaka od ovih paradigmi ima različite karakteristike u smislu (a) brzine, (b) snage, (c) veličine, (d) potencijala za visoku preciznost i lakoće programiranja. Svaka od njih je najpogodnija za određeni skup problema. Neke paradigme su pogodnije da služe kao host računari, a druge kao akceleratori. Naravno, najbolje ih je iskoristiti kroz odgovarajuću vrstu sinergije.

Ovaj članak prvo predstavlja prednosti i nedostatke svakog od njih, a zatim raspravlja o mogućim načinima za njihovu sinergiju.

2. POREĐENJE ČETIRI RAČUNARSKE PARADIGME

Control Flow paradigma zasniva se na istraživanju Fon Nojmana. Ona je pogodna za transakcijsko računarstvo i mogla bi se efikasno koristiti kao host u hibridnim mašinama, koje kombinuju sve gore pomenute paradigme. U slučaju kada se Control Flow MultiCore mašina koristi kao host, transakcioni kôd je najbolje pokrenuti na Control Flow hostu, dok se druge vrste problema najbolje rešavaju na akceleratorima zasnovanim na drugim tipovima paradigmi. U slučaju kada kôd radi na podacima organizovanim u 2D, 3D ili n-dimenzionalnim strukturama, dobar nivo ubrzanja može se postići pomoću Control Flowa ManyCore akceleratora. Model programiranja je relativno lak za razumevanje. Brzina, snaga, veličina i potencijal za visoku preciznost mašina Control Flow su dobro poznati.

Data Flow paradigma inspirisana je istraživanjem Ričarda Fejnmana i drugih i insistira na činjenici da je računarstvo najefikasnije ako se podaci prenose, tokom računarskog procesa, na beskonačno male udaljenosti, kao u slučaju računarstva zasnovanog na grafu izvršenja (executon graph). U poređenju sa Control Flowom, ovaj pristup donosi ubrzanja, uštedu energije, manje dimenzije računarske opreme i veći potencijal za povećanu preciznost, ali koristi složeniji model programiranja, koji bi se mogao podići na više nivo apstrakcije, u kom slučaju bi deo navedenih prednosti mogao nestati.

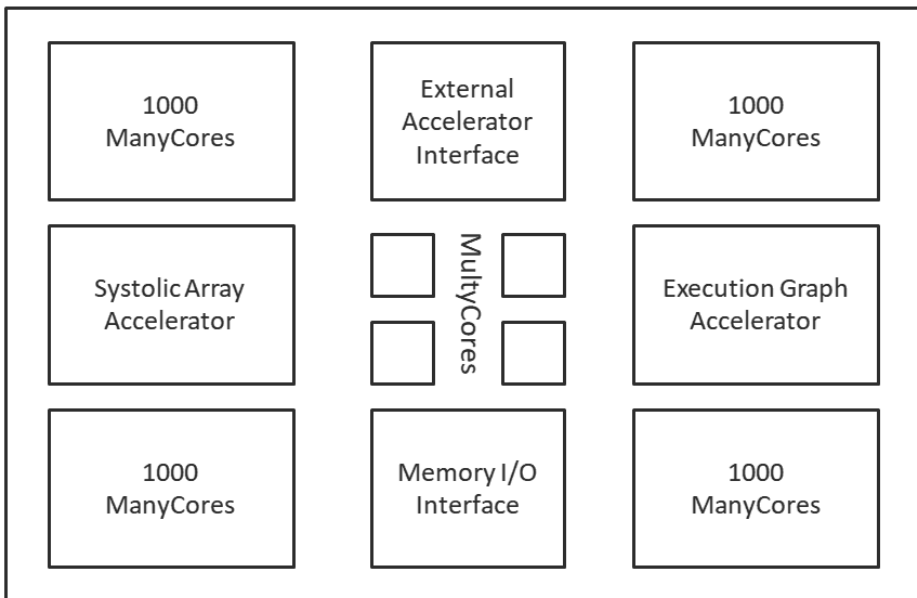
Diffusion Flow paradigma zasniva se na istraživanju masivnog paralelizma (IoT) moguće poboljšanog senzorskim mrežama (VSN). Jedna suštinska karakteristika ovog pristupa je velika oblast ili geografska pokrivenost, što znači da je teoretski nemoguće preneti podatke na male udaljenosti tokom računarske obrade. Ipak, neki nivo obrade je neophodan, možda u svrhu smanjenja podataka, ili neke vrste predobrade tokom „difuzije“ prikupljenih podataka prema hostu, za konačnu obradu tipa Big Data. Energetska efikasnost je visoka, dok je veličina zanemarljiva, kao i potencijal za najveću

preciznost. S druge strane, model programiranja je evoluirao od početnog MIT PROTO pristupa i mora se pravilno savladati, što nije jednostavno.

Paradigma protoka energije Energy Flowa je namenjena samo za ubrzanje algoritama koji su najpogodniji za jednu od postojećih paradigmi. Bez obzira da li se koristi biomolekularni ili kvantnomehanički pristup, obrada se zasniva na energetskim transformacijama, a odgovarajući model programiranja mora da poštuje unutrašnju suštinu ako su žele najveće performanse. Za izvodljive algoritme ubrzanje je ogromno, potrebna snaga je minimalna, veličina je prihvatljiva, a potencijal za preciznost je nezamislivo veliki. Modeli programiranja su u porastu.

3. MOGUĆA ARHITEKTURA SUPERKOMPJUTERA NA ČIPU

U trenutnom stanju tehnologije, sa preko 100 milijardi tranzistora (BTr) na čipu, ili trilionu tranzistora (TTr) na pločici (wafer), moguće je postaviti (na jedan čip) obe gore pomenute Control Flow mašine i obe gore pomenute Data Flow mašine. Međutim, mogući pojačivači (u obliku IoT-a ili VSN-a) i mogući akceleratori (u obliku biomolekularnog i/ili kvantnomehanskog) moraju biti van čipa, ali lako dostupni preko odgovarajućih interfejsa.



Slika 1. Generička struktura budućeg superračunara na čipu sa 100 milijardi tranzistora [10]

Naravno, memorija i klasični I/O moraju biti delimično na čipu, a delimično van čipa, povezivi odgovarajućim interfejsima.

Stoga, bez obzira da li se radi o 100 BTr ili 1 TTr strukturama, unutrašnja arhitektura na najvišem nivou apstrakcije treba da bude kao na Slici 1. Međutim, raspodela resursa bi mogla da se drastično razlikuje od jednog takvog čipa do drugog, zbog različitih zahteva aplikacija (orijentisani na transakcije ili orijentisani na „krckanje podataka“ — transactions-oriented or cranching-oriented), i zbog različitih zahteva za podacima (intenzivni zahtevi za memorijom zbog Big Data (masivnih podataka) statičkog tipa, ili intenzivni zahtevi za striming (prolazno) orijentisanih masivnih podataka dinamičkog tipa, koji dolaze i odlaze preko Interneta ili drugih protokola. Primeri koji slede pokrivaju simulacije problema velikih podataka (Big Data), gde je potrebno mašinsko učenje i odnose se na složene probleme u građevinarstvu ili srodnim oblastima:

- (1) NBCE — građevinski inženjering zasnovan na prirodi;
- (2) GNBE — genomska podrška za NBCE;
- (3) EKIS — informacioni sistemi zemljotresa za predviđanje i alarm;
- (4) NCEM — izrada novih građevinskih materijala za CO₂ i zemljotrese.

Algoritam koji se koristi u gore navedenim oblastima može biti:

- (1) statistički i slučajni procesi koji oponašaju prirodu;
- (2) NW ili SW⁵ ili slično;
- (3) PDE⁶ tipa FE ili FD ili hibrid (FE = konačni element, FD = konačna razlika);
- (4) tenzorski račun i matematička logika ili hibrid.

Za takav skup aplikacija pretpostavljamo da bi optimalna raspodela resursa bila kao u Tabeli 1.

4. NAPOMENA

U NBCE-u je bolje koristiti biološke strukture koje brzo rastu i naseljene su insektima koji stvaraju nanomaterijale, nego graditi betonske zidove, koji emituju CO₂ i koji su osetljivi na zemljotrese. Takođe, za zaštitu podvodnih objekata bolje je koristiti ribu i plankton nego metalne mreže. Pre svake investicije ove vrste, mora se uraditi studija izvodljivosti, zasnovana na simulaciji. Međutim, takve simulacije mogu biti dugotrajne i mogu trajati

⁵ NW — Needleman Wunsch; SW — Smith Waterman

⁶ PDE — Partial Differential Equations — parcijalne diferencijalne jednačine

Tabela 1. Tip hardvera čipa i procenjeni broj tranzistora [10]

Tip Hardvera čipa	Procenjeni broj tranzistora
Jedan ManyCore sa memorijom	3,29 miliona
Jedan sistolni niz	< 1 milijarde
4000 ManyCore sa memorijom	11 800 miliona
Jedan vrhunski Data Flow, koji se može ponovno reprogramirati	< 69 milijardi
Jedan MultiCore sa memorijom	1 milijarda
Interfejs za I/O sa spoljnom memorijom	< 100 miliona
4 MultiCore sa memorijom	4 milijarde
Interfejs za spoljne akceleratora	< 100 miliona
TOTAL	< 100 milijardi

godinama. Rešenje je u prelasku sa Control Flowa na odgovarajuću kombinaciju ostale tri računarske paradigme.

U GNBE-u genetika vrsta i srodni procesi mogu u stvarnosti proći godine da se generišu željeni efekti za građevinarstvo. Međutim, kompjuterske simulacije na Control Flow mašinama, zasnovane na dovoljno detalja, mogle bi da potraju još više vremena. Dakle, rešenje je u odgovarajućoj sinergiji četiri paradigme.

U EKIS-u postoje modeli gradova zasnovani na cigli i cementu, ali simulacije zemljotresa sa ovim modelima kao ulazima mogu potrajati vek na najbržoj Control Flow mašini danas. Proces simulacije može se drastično ubrzati samo ako se koristi odgovarajući akcelerator protoka podataka. Pogodni su za PDE-ove tipa FE (Final Element), koji su potrebni za predviđanja i za PDE-ove tipa FD (Final Difference), koji su potrebni za alarmiranje u hitnim slučajevima.

U NCEM-u, novi materijali sa željenim svojstvima se najbolje nalaze ako se ML (Machine Learning — mašinsko učenje) algoritmi kombinuju sa klasičnim algoritmima, koji se koriste u istraživanju materijala. Ovakvi hibridni algoritmi su računarski intenzivni, pa je opet rešenje u sinergiji nekoliko paradigmi.

5. ZAKLJUČAK

Ovaj članak baca svetlo na potencijale koji dolaze iz sinergijskih interakcija četiri različite računarske paradigme.

Ovde primenjen pristup se razmatra u kontekstu građevinarstva, ali se lako može preneti u druge različite kontekste.

Test ovog tipa mogao bi se koristiti u obrazovne ili istraživačke svrhe u akademskoj zajednici i industriji.

Koncept koji se zagovara u ovom članku najbolje je primeniti na čipu koji uključuje neke od paradigmi koje se češće koriste i koji se efikasno povezuje sa drugim paradigrama koje se ređe koriste!

BIBLIOGRAFIJA

- [1] J. L. Henning, „SPEC CPU2000: measuring CPU performance in the New Millennium,“ in *Computer*, vol. 33, no. 7, pp. 28–35, July 2000, DOI: 10.1109/2.869367.
- [2] Mittal, Sparsh and Vetter, Jeffrey S, „A Survey of CPU-GPU Heterogeneous Computing Techniques“, in *ACM Computing Surveys*, Volume 47, Issue 4, July 2015, Article No.: 69, pp. 1–35. <https://doi.org/10.1145/2788396>
- [3] Kumar, Sameer et al., „Scale MLPerf-0.6 models on Google TPU-v3 Pods“, in *Computer Science, Cornell University*, <https://doi.org/10.48550/arXiv.1909.09756>
- [4] Wang, Yu Emma; Wei, Gu-Yeon and Brooksar, David, „Benchmarking TPU, GPU, and CPU Platforms for Deep Learning“, in *Computer Science, Cornell University*, Oct 2019. <https://doi.org/10.48550/arXiv.1907.10701>
- [5] Trifunovic, N.; Milutinovic, V.; Salom, J. et al., „Paradigm Shift in Big Data SuperComputing: DataFlow vs. ControlFlow“, in *Journal of Big Data* 2, Springer, 4 (2015). <https://doi.org/10.1186/s40537-014-0010-z>
- [6] Srivastava, Ankita and Mishra, Pramod Kumar, „A Survey on WSN Issues with its Heuristics and Meta-Heuristics Solutions“, in *Wireless Personal Communication* 121, pp. 745–814 (2021). <https://doi.org/10.1007/s11277-021-08659-x>
- [7] M. Centenaro; C. E. Costa; F. Granelli; C. Sacchi and L. Vangelista, „A Survey on Technologies, Standards and Open Challenges in Satellite IoT“, in *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1693–1720, 2021, DOI: 10.1109/COMST.2021.3078433.
- [8] Schlick T.; Portillo-Ledesma S.; Myers C. G. et al., „Biomolecular Modeling and Simulation: A Prospering Multidisciplinary Field“, in *Annual Review of Biophysics*. 2021 May; 50: 267–301. DOI: 10.1146/annurev-biophys-091720-102019. PMID: 33606945; PMCID: PMC8105287.
- [9] Baiardi, Alberto; Grimm, Stephanie A.; Steiner, Miguel; et al., „Expansive Quantum Mechanical Exploration of Chemical Reaction Paths“, *Laboratory of Physical Chemistry, ETH Zurich, Acc. Chem. Res.* 2022, 55, 1, 35–43 Publication Date: December 17, 2021. <https://doi.org/10.1021/acs.accounts.1c00472>[10]
- [10] Milutinovic, Veljko; Azer, Erfan Sadeqi; Yoshimoto, Kristy et al., „The Ultimate DataFlow for Ultimate SuperComputers-on-a-Chip, for Scientific Computing, Geo Physics, Complex Mathematics, and Information Processing“, 10th Mediterranean Conference on Embedded Computing (MECO), (2021), pp. 1–6, DOI: 10.1109/MECO52532.2021.9459725.