

Studi Kasus Evaluasi Batasan Dasar Komputasi

Suwarno

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Internasional
Batam, Jln Gajah Mada No 1, 29432, Indonesia

E-mail: suwarno.liang@uib.ac.id

Abstrak

Komputasi, sebagai bagian tak terpisahkan dari kehidupan pribadi dan kerja kita, juga menjadi penting bagi industri dan pemerintah. Perbaikan perangkat keras komputer yang stabil telah didukung oleh peningkatan dua kali lipat kerapatan transistor di sirkuit terpadu selama lima puluh tahun terakhir. Skala penutur Moore semacam itu membutuhkan usaha yang terus meningkat, merangsang penelitian di perangkat keras alternatif dan menimbulkan kontroversi. Untuk membantu mengevaluasi teknologi yang muncul dan meningkatkan pemahaman kita, maka dari skala sirkuit terpadu, di sini peneliti meninjau batasan mendasar untuk perhitungan di bidang manufaktur, energi, ruang fisik, desain dan upaya verifikasi, dan algoritma. Untuk menjelaskan apa yang dapat dicapai secara prinsip dan dalam praktiknya, peneliti rekapitulasi bagaimana beberapa batasan dielakkan, dan bandingkan batasan longgar dan ketat. Kesulitan teknik yang dihadapi oleh teknologi baru mungkin mengindikasikan batasan yang belum diketahui.

As an indispensable part of our personal and working lives, computing has also become essential to industries and governments. Steady improvements in computer hardware have been supported by periodic doubling of transistor densities in integrated circuits over the past fifty years. Such Moore scaling now requires ever increasing efforts, stimulating research in alternative hardware and stirring controversy. To help evaluate emerging technologies and increase our understanding of integrated-circuit scaling, here I review fundamental limits to computation in the areas of manufacturing, energy, physical space, design and verification effort, and algorithms. To outline what is achievable in principle and in practice, I recapitulate how some limits were circumvented, and compare loose and tight limits. Engineering difficulties encountered by emerging technologies may indicate yet unknown limits.

Katakunci: komputasi, evaluasi, batasan, dasar

1. Latar Belakang

Teknologi baru muncul untuk komputasi menjanjikan untuk mengungguli konvensional, dimana cara-cara yang lama belum dapat memenuhi kebutuhan

kecepatan dari aplikasi dan basis data yang selama ini terus meningkat, seiring dengan perkembangan dari multimedia dan berbagai aplikasi chatting seperti Skype, WhatsApp, atau Messenger dari Facebook. Aplikasi-aplikasi macam ini

mempertimbangkan universalitas) [3]. Efisiensi dipelajari dari perspektif teoritis pada awalnya, namun permintaan yang kuat dalam aplikasi militer pada tahun 1940an membuat Turing dan von Neumann mengembangkan arsitektur perangkat keras yang terperinci untuk komputer universal-desain Turing (PilotACE) lebih efisien, tapi program von Neumann lebih mudah diprogram.

Arsitektur program tersimpan membuat komputer universal praktis dalam arti bahwa satu desain komputer bisa menjadi efektif dalam beragam aplikasi jika disertakan dengan perangkat lunak yang sesuai. Universalitas praktis semacam itu tumbuh subur (1) dalam skala ekonomi di perangkat keras komputer dan (2) di antara tumpukan perangkat lunak yang ekstensif. Tidak mengherankan, desain dan komponen komputer yang paling canggih dan sukses secara komersial, seperti unit pemrosesan IBM Intel dan IBM (CPU), didasarkan pada paradigma von Neumann.

Beragam penggunaan dan pasar besar chip tujuan umum, serta kemampuan reproduksi hasilnya yang tepat, membenarkan investasi modal yang sangat besar dalam perancangan, verifikasi dan pembuatan sirkuit terintegrasi terdepan. Hari ini tujuan umum CPU daya server awan peternakan dan menggantikan prosesor mainframe khusus (tapi masih universal) di banyak superkomputer. Komputer universal yang berkembang berdasarkan susunan gerbang programmable yang dapat diprogram dan unit pemrosesan grafis tujuan umum mengungguli CPU dalam beberapa kasus, namun efisiensinya tetap saling melengkapi. Untuk CPU. Keberhasilan komputasi deterministik tujuan umum Terwujud dalam konvergensi beragam fungsi pada smartphone portabel dan murah. Setelah perbaikan mantap, komputasi tujuan umum

2. Universal and general-purpose computers

Jika kita melihat jam dan jam tangan sebagai komputer awal, mudah untuk melihat pentingnya perhitungan lama yang dapat diulang dengan akurasi tinggi oleh perangkat yang diproduksi secara massal. Pentingnya komputer digital yang diprogram menjadi jelas setidaknya 200 tahun yang lalu, seperti yang diilustrasikan oleh alat tenun Jacquard di manufaktur tekstil. Namun, keberadaan komputer universal yang dapat secara efisien mensimulasikan (hampir) semua perangkat komputasi lainnya - analog atau digital - hanya diartikulasikan pada tahun 1930an oleh Gereja dan Turing (Turing mengecualikan kuantum saat

Seluruh industri yang terlantar (surat kabar, fotografi, dan sebagainya) Dan meluncurkan aplikasi baru (konferensi video, navigasi GPS, online Belanja, hiburan berjejaring, dan sebagainya) [4]. Aplikasi spesifik terintegrasi sirkuit merampingkan input-output dan networking, atau mengoptimalkan fungsionalitas Sebelumnya dilakukan oleh hardware general purpose. Kecepatan mereka Up simulasi biomolekuler 100 kali lipat [5], [6] dan meningkatkan efisiensi video decoding 500 kali lipat [7], namun membutuhkan usaha perancangan dengan pemahaman yang tajam Dari perhitungan khusus, memberlakukan biaya tinggi dan risiko finansial, perlu pasar swhere Komputer tujuan umum ketinggalan, dan seringkali tidak dapat beradaptasi dengan algoritma baru. Teknik terbaru untuk komputasi spesifik domain yang dapat disesuaikan [8] menawarkan pengorbanan yang lebih baik, sementara banyak aplikasi menyukai kombinasi perangkat keras tujuan umum dan perangkat lunak khusus domain, termasuk bahasa pemrograman khusus [9], [10] seperti Erlang, yang digunakan Untuk mengimplementasikan instant messenger Whatsapp yang populer.

3. Batasan Sebagai Alat Bantu Untuk Mengevaluasi Teknologi Baru

Tanpa sejarah yang memadai, kita tidak bisa melakukan ekstrapolasi hukum penskalaan untuk teknologi baru, namun ekspektasi berjalan tinggi. Sebagai contoh, proposal baru untuk prosesor analog sering muncul (seperti yang diilustrasikan oleh kuantumkomputer adiabatik), namun gagal mengatasi masalah tentang komputasi analog, seperti keterbatasan pada skala, keandalan, dan perhitungan bebas-

kesalahan yang berjalan lama. Komputer tujuan umum memenuhi persyaratan ini dengan sirkuit terpadu digital dan sekarang menguasai pasar elektronik. Sebagai perbandingan, komputer kuantum-baik janji digital maupun analog-hold hanya di aplikasi niche dan tidak menawarkan komputasi tujuan umum yang lebih cepat karena mereka tidak cepat untuk menyortir dan tugas spesifik lainnya. Dalam melebihi-lebihkan dampak rekayasa dari komputer kuantum, pers yang populer telah melewati poin penting ini. Namun dalam penelitian ilmiah, upaya untuk membangun komputer kuantum memungkinkan dalam mensimulasikan fenomena kimia kuantum dan mengungkapkan batasan fundamental baru. Bagian 'Batas waktu-waktu asimtotik' dan 'Kesimpulan' mendeskripsikan batasan teknologi yang muncul.

Ekstrapolasi teknologi versus batas fundamental

Penskalaan perangkat keras komputasi komersial secara teratur mengalami hambatan yang luar biasa [1] [2], namun kemajuan teknologi jangka pendek seringkali menjauhi mereka.

Sebagai contoh, perkiraan konsensus pada awalnya memperkirakan 10-GHz CPU untuk simpul 45 nmteknologi 15, versus rentang 3-4 GHz yang terlihat dalam praktik. Pada tahun 2004, Peta Jalan Ilmu Pengetahuan dan Teknologi Informasi Quantum yang tidak terkait memperkirakan 50 qubit fisik 'digital' pada tahun 2012. Optimisme semacam itu muncul dengan mengasumsikan solusi teknologi jauh sebelum dikembangkan dan divalidasi, dan dengan mengabaikan batasan penting. Penulis mengklasifikasikan batas ke perangkat dan interkoneksi sebagai rangkaian, materi, perangkat, sirkuit, dan sistem yang mendasar. Batas spesifik

pada teknologi komputasi yang layak, mencari batasan 'ketat', yang menghalangi perbaikan parameter kunci jangka.

4. Batasan Daya dan CPU

Pada tahun 2004, Intel secara tiba-tiba membatalkan proyek CPU 4-GHz karena kerapatan daya yang tinggi memerlukan teknologi pendinginan yang kikuk. Pabrik CPU lainnya menyimpan frekuensi clock di 1-6-GHz range. Daya sirkuit sinedynamic tumbuh dengan frekuensi clock dan voltase suplai kuadrat [11], energi dapat diselamatkan dengan cara mendistribusikan pekerjaan di antara yang lebih lambat, voltase rendah.

Inti CPU paralel jika overhead paralelisasi kecil.

Tren pendamping terhadap hukum Moore - teori penskalaan Dennard [12] - menunjukkan bagaimana menjaga konsumsi daya dari sirkuit terpadu semikonduktor tetap konstan sambil meningkatkan kerapatannya. Tapi penskalaan Dennard mogok sepuluh tahun yang lalu [12]. Ekstrapolasi tren penskalaan semikonduktor untuk CMOSs - teknologi semikonduktor yang dominan selama 20 tahun terakhir - menunjukkan bahwa konsumsi daya transistor yang tersedia di sirkuit terpadu modern berkurang lebih lambat dari ukurannya (yang tunduk pada hukum China) [13], [14]. Untuk memastikan karakteristik kinerja transistor yang dapat diterima, kerapatan daya chip harus dibatasi, dan pecahan transistor harus tetap gelap pada suatu waktu. CPU modern belum dapat menggunakan semua sirkuit mereka sekaligus, namun efek asimtotik ini disebut "Dinding pemanfaatan"49-akan segera mengeluarkan 99% chip, maka istilah 'silikon gelap' dan referensi beralasan

pada kiamat [13]. Hemat tenaga dengan memperlambat core CPU turun disebut 'silicon redup'. Studi terperinci mengenai hasil silikon [14] yang gelap menunjukkan hasil yang sama. Untuk tujuan ini, eksekutif dari Microsoft dan IBM baru-baru ini memproklamasikan akhir era multicoremicroprocessors [15]. Dua tren terkait muncul sebelumnya:

(1) daerah sirkuit terpadu yang semakin besar tetap bebas dari transistor untuk membantu rute dan sintesis fisik, untuk mengakomodasi jaringan pasokan daya, dan seterusnya [16], [17] - kita menyebutnya 'silikon yang lebih gelap', (2) semakin banyak gerbang tidak berfungsi Perhitungan tapi memperkuat interkoneksi panjang, lemah [18] atau memperlambat kabel yang terlalu pendek - yang peneliti sebut 'silikon abu-abu'.

Saat ini, 50% -80% dari semua gerbang di sirkuit terpadu berkinerja tinggi adalah repeater. Batas untuk catu daya dan pendinginan. Pusat data di USA menguji 2,2% dari total listrik di 2011. Karena pembangkit listrik memerlukan waktu untuk membangun, kita tidak dapat mempertahankan tren konsumsi daya dua kali lipat per tahun. Hal ini dimungkinkan untuk meningkatkan efisiensi jalur transmisi (menggunakan superkonduktor hightempergi [19]) dan konversi daya di pusat data, namun efisiensi jaringan daya on-chip dapat segera mencapai 80% -90%, sehingga tidak memberi banyak ruang untuk perbaikan. Manajemen daya sirkuit terpadu yang modern mencakup jaringan jam dan tenaga gating, tegangan per-core scaling [20], pemulihan biaya [21] dan, pada prosesor terbaru, inti CPU yang didedikasikan untuk penjadwalan daya. Konsumsi daya sirkuit terpadu bergantung secara kuadrat pada voltase suplai, yang telah menurun dengan mantap selama bertahun-tahun, namun baru-baru ini stabil pada 0,5-2V. Tegangan suplai biasanya melebihi

voltase ambang batas dengan margin keamanan yang menjamin keandalan sirkuit, operasi cepat dan kebocoran yang rendah. Tegangan ambang batas bergantung pada ketebalan gerbang dielektrik, yang mencapai batas praktis beberapa atom (lihat bagian 'Hambatan Rekayasa') . Transistor tidak dapat beroperasi dengan voltase suplai di bawah sekitar 200mV -lima kali di bawah praktik saat ini-dan rangkaian sederhana mencapai batas ini. Dengan operasi yang lebih lambat, sirkuit di dekat dan sub-ambang batas menghabiskan energi seratus kali lebih sedikit [22]. Teknologi pendingin juga bisa meningkat, namun batas kuantum mendasar membatasi efisiensi penghilangan panas [23]

5. Teori Limit Kompleksitas

Sekarang kita tinjau model yang lebih abstrak untuk fokus pada dampak skala, dan untuk menunjukkan bagaimana tren berulang dengan cepat menyialip efek spesifik perangkat satu kali. Peneliti mengabaikan efek spasial dan fokus pada sifat perhitungan dalam model abstrak (digunakan oleh insinyur perangkat lunak) yang merupakan perhitungan dengan langkah-langkah elementer dengan runtimes input-independen. Batas tersebut bertahan banyak perbaikan dalam teknologi komputer, dan seringkali lebih kuat untuk masalah tertentu. Sebagai contoh, algoritma yang paling terkenal untuk mengalikan bilangan besar hanya sedikit lebih lambat daripada membaca input (batas kecepatan yang jelas), namun hanya dalam pengertian asimtotik: untuk angka dengan kurang dari seribu bit, algoritma tersebut tertinggal dari algoritma yang lebih sederhana dibandingkan kinerja sebenarnya. Untuk memusatkan perhatian pada hal yang paling penting, saya tidak lagi melacak kompleksitas kasus asymptotic

terburuk dari algoritma terbaik untuk masalah tertentu, namun hanya membedakan pertumbuhan asimtotik polinomial dari eksponensial.

Batasan yang diformulasikan dengan persyaratan kasar (unsolvability in polinomial time on any computer) sangat kuat: kekerasan anjak piutang mendasari perdagangan Internet, sementara P tidak sama dengan dugaan NP menjelaskan kurangnya solusi terukur dan terukur untuk masalah algoritmik yang penting. , Dalam optimasi dan verifikasi desain sirkuit terpadu, misalnya. (Disini P adalah kelas masalah keputusan yang dapat dipecahkan dengan menggunakan langkah-langkah perhitungan sederhana yang jumlahnya tidak lebih cepat dari pada polinomial dengan ukuran data masukan, dan NP adalah Kelas polinomial non-deterministik yang mewakili masalah keputusan yang solusi yang dapat ditentukan secara deterministik dapat diperiksa dengan tepat dengan menggunakan sejumlah langkah polinomial.) Asimilar yang sama, P tidak sama dengan NC, berusaha untuk menjelaskan mengapa banyak masalah algoritmik yang dapat dipecahkan Secara efisien tidak paralel secara efisien. Sebagian besar batasan ini belum terbukti. Beberapa dapat dielakkan dengan menggunakan fisika yang sangat berbeda, misalnya, komputer kuantum dapat memecahkan anjuran bilangan pada waktu polinomial (secara teori). Namun perhitungan kuantum tidak mempengaruhi P tidak sama dengan NP. Kurangnya bukti, meski bukti empiris berat, membutuhkan kepercayaan dan merupakan batasan penting dari banyak batasan nonfisik untuk komputasi. Arare membuktikan hasil oleh Turing menyatakan bahwa memeriksa apakah suatu program yang pernah dihentikan tidak dapat dipecahkan: tidak ada algoritma yang memecahkan masalah ini dalam semua kasus tanpa mempedulikan runtime.

Namun, pengembang perangkat lunak memecahkan masalah ini selama ulasan kode peer, dan begitu juga guru sains komputer saat mengerjakan ujian dalam kursus pemrograman.

Analisis kasus terburuk adalah batasan batasan nonfisik lainnya untuk komputasi, namun menunjukkan potensi keuntungan melalui pendekatan dan spesialisasi.

Untuk beberapa masalah optimasi NP-hard, seperti Soal Traveling Salesman Euclidean, perkiraan waktu polinomial ada, namun dalam kasus lain, seperti masalah Maksimum Clique, perkiraan yang akurat sama sulitnya dengan menemukan solusi optimal. Untuk beberapa masalah penting dan Algoritma, seperti algoritma Simplex untuk pemrograman linier, beberapa masukan mengarah pada eksponensial runtime, dan gangguan menit mengurangi runtime ke polinom.

6. Kesimpulan

Pawai kematian hukum Moore mengundang diskusi tentang batasan fundamental dan alternatif semikonduktor silikon. Kendala jangka pendek (hambatan terhadap kinerja, kekuatan, bahan, sumber laser, teknologi manufaktur dan sebagainya) selalu terkait dengan biaya dan modal, namun diabaikan karena adanya pasar baru untuk elektronik terbuka, populasi meningkat, dan ekonomi dunia tumbuh. Tekanan Suheconomic menekankan nilai universalitas komputasi dan penerapan arsitektur sirkuit terpadu yang luas untuk menyelesaikan banyak tugas di bawah kondisi lingkungan konvensional. Dalam skenario yang mungkin, hanya CPU, unit pemrosesan grafis, susunan gerbang yang dapat diprogram di lapangan dan sirkuit terpadu padat perhatian akan tetap bertahan pada akhir hukum Moore,

sementara sirkuit khusus akan didominasi oleh teknologi yang kurang canggih karena alasan keuangan. Mengam, chip memori telah mencontohkan penskalaan Moore karena strukturnya yang sederhana, interkoneksi sederhana, dan manufaktur yang lebih terkendali, namun miniaturisasi sel memori sekarang melambat. Penipisan sirkuit CMOS yang melambat masih mengguguli penskalaan teknologi baru yang paling layak.

Referensi

- [1 R. Cavin, P. Lugli and V. Zhimov,] "Science and Engineering Beyond Moore's Law," 2012.
- [2 A. Chien and V. Karamcheti,] "Moore's Law: the First Ending and New Beginning," *IEEE Computer*, pp. 48-53, 2013.
- [3 R. Herken, The Universal Turing] Machine: A Half Century Survey 2nd edn, Springer, 2013.
- [4 M. Andreessen, "Why Software is] eating the world," *The Wall Street Journal*, 2011.
- [5 D. Padua, Encyclopedia of Pararell] Computing, Springer, 2011.
- [6 D. A. Shaw, "A Special Purpose] Machine that Achieves a Hundred-Fold Speedup in Biomolecular

- Simulations," *IEEE*, pp. 129-130, 2013.
- [7 R. e. a. Hameed, "Understanding Sources of Inefficiency in General Purpose Chips," *Communication ACM*, pp. 85-93, 2011.
- [8 J. Cong, G. Reinman, A. Bui and V. Sarkar, "Customizable Domain-Specific Computing," *IEEE*, pp. 6-15, 2011.
- [9 M. Mernik, J. Heering and A. Sloane, "When and How to Develop Domain-Specific Languages," *ACM Computing Surv.*, pp. 316-344, 2005.
- [1 K. Olukotun, Beyond Pararell 0] Programming With Domain Specific Languages, ACM, 2014, pp. 179-180.
- [1 J. Rabaey, A. Chandrakasan and B. Nikolic, Digital Integrated Circuits A Design Perspective, Pearson Education, 2003, pp. 11-13.
- [1 M. Bohr, "A 30 Year retrospective on 2] Dennard's MOSFET Scaling Paper," *IEEE*, pp. 11-13, 2007.
- [1 M. Taylor, "Is Dark Silicon Useful? 3] Harnessing the Four Horsemen of the Coming Dark Silicon Apocalypse," *ACM*, pp. 1131-1136, 2012.
- [1 H. Esmailzadeh, E. Blem, R. St-4] Amant, K. Sankaralingam and D. Burger, "Power Challenges May End the Multicore Era," *ACM*, pp. 93-102, 2013.
- [1 Z. Yeraswork, "3D Stacks and 5] Security Key for IBM in Server Market," 17 December 2013. [Online]. Available: http://www.eetimes.com/document.asp?doc_id=1320403.
- [1 A. Caldwell, A. Kahng and I. 6] Markov, "Hierarchical Whitespace Allocation in Top-Down Placement," *IEEE*, pp. 716-724, 2003.
- [1 S. Adya, I. Markov and P. 7] Villarubia, "On Whitespace and Stability in Physical Synthesis," *Integration*, pp. 340-362, 2006.
- [1 P. Saxena, N. Menezes, P. 8] Cocchini and D. Kirkpatrick, "Repeater Scaling and Its Impact on CAD," *IEEE*, pp. 451-463, 2004.
- [1 J. Oestergaard, J. OKholm, K. 9] LOMholt and G. Toennesen, "Energy Losses of Superconducting Power Transmission Cables in the Grid," *IEEE*, p. 11, 2001.
- [2 N. E. a. Pinckney, "Limits of 0] Pararellism and Boosting in Dim Silicon," *IEEE*, pp. 30-37, 2013.

[2 S. Kim, Z. C.H. and M.

- 1) Papaefthymiou, "Charge-Recovery Computing on Silicon," *IEEE*, pp. 651-659, 2005.

[2 R. Drelinski, M. Wieckowski, D.

- 2) Blaauw, D. Sylvester and T. Mudge, "Near-threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits," *IEEE*, pp. 253-266, 2010.

[2 J. Pendry, "Quantum Limits to The

- 3) Flow of Information and Entropy," *Physic and Mathematics Gen.*, pp. 2161-2171, 1983.