Modelle der Parallelverarbeitung

10. «Realistische» parallele Modelle

Thomas Worsch

Institut für Theoretische Informatik
Karlsruher Institut für Technologie

Sommersemester 2018

Überblick

Physikalisch realisierbare Schaltkreisfamilien

Zellularautomaten mit realistischer «Abwärme»

Was ist «physikalisch realisierbar»?

Was ist «physikalisch realisierbar»?

- Nur endliche Automaten?
  - selbst wenn: Ist das immer der angemessene Blickpunkt?
- allgemeiner: Wieviele Verarbeitungseinheiten stehen für Eingaben der Länge n zur Verfügung?
- Wie lang dürfen Drähte sein?
- Was ist mit Wärmeentwicklung?

Überblick

Physikalisch realisierbare Schaltkreisfamilien

Zellularautomaten mit realistischer «Abwärme»

Literatur

A. R. Schorr
Physical parallel devices are not much faster than sequential ones.
Information Processing Letters, Band 17, 103–106, 1983.

Annahmen

- dreidimensionale euklidische Welt
- deterministische digitale Berechnungen mit Schaltkreisen
- Gatter mit Volumen $v > 0$
- Informationstransport über Entfernung d benötigt mindestens Zeit d/c (c Lichtgeschwindigkeit)

Physikalisch realisierbare Schaltkreisfamilien

- genügen den eben gemachten Annahmen
- Statt Tiefe Depth_R betrachte man «Zeitbedarf» Time_R, der durch die Datenübertragungszeiten dominiert wird.

Folgerungen

Es sei $R = (R_n)_{n \in \mathbb{N}_+}$ eine physikalisch realisierbare Schaltkreisfamilie.
Dann:

- Für die Volumina $\mathrm{Vol}_R(n)$ der R_n gilt:
  Für alle $n \in \mathbb{N}_+$: $\mathrm{Vol}_R(n) \geq v \cdot \mathrm{Size}_R(n)$.
- Es bezeichne $l_R(n)$ den maximalen geometrischen Abstand eines Gatters vom Ausgang in R_n.
  Dann existiert eine Konstante a, so dass für alle $n \in \mathbb{N}_+$ gilt:
  $l_R^3(n) \geq a \cdot \mathrm{Size}_R(n)$.

Lemma

Es sei L eine formale Sprache und R eine physikalisch realisierbare Schaltkreisfamilie, die L erkennt und unter allen solchen Familien die kleinste Komplexität Size_R hat. Dann gibt es eine Konstante a' mit: $\mathrm{Time}_R^3 \geq a' \cdot \mathrm{Size}_R$.

Beachte:

- Minimalitätsforderung ok, da keine Uniformität verlangt

Folgerungen:

- $\mathrm{Time}_R \geq a' {\mathrm{Size}_R}^{1/3}$
- Für jedes nichttriviale Problem ist $\mathrm{Size}_R \in \Omega(n)$ und also *z. B. logarithmische Zeit unmöglich*.

Beweis des Lemmas

- Es sei R wie gefordert.
- Von jedem Gatter muss Information zum Ausgang transportiert werden.
- Zeitbedarf dafür $\geq l_R/c$.
- Also $\mathrm{Time}_R \geq l_R/c$.
- Also $\mathrm{Time}_R^3 \geq l_R^3/c^3 \geq a/c^3 \cdot \mathrm{Size}_R$.

Satz

Ist U eine uniforme, physikalisch realisierbare Schaltkreisfamilie, dann gibt es ein Polynom p und eine Turingmaschine T, die die gleiche Sprache $L(U)$ erkennt und für deren Zeitkomplexität gilt: $\mathrm{Time}_T \leq p(\mathrm{Time}_U)$.

Beweis

- Es sei U eine uniforme, physikalisch realisierbare Schaltkreisfamilie.
- Dann gibt es eine äquivalente Turingmaschine T und ein Polynom p' mit $\mathrm{Time}_T \leq p'(\mathrm{Size}_U)$.
- Nach vorangegangenen Lemma ist $\mathrm{Size}_U \leq p''(\mathrm{Time}_U)$ für ein Polynom p''.
- also $\mathrm{Time}_T \leq p'(p''(\mathrm{Time}_U))$
- also existiert Polynom p wie gewünscht

Überblick

Überblick

Annahmen

- physikalische Grenzen:
  - Übertragung von Information benötigt Zeit (proportional zur Entfernung)
  - Verarbeitungseinheiten haben konstant viele Eingänge und endliches, nicht verschwindendes Volumen.
  - euklidische r-dimensionale Geometrie der Welt
- technologische Einschränkungen:
  - klassische, deterministische, digitale Berechnungen
  - Berechnungen verbrauchen Energie
    - *wird gleich noch diskutiert*
  - Es gibt eine obere Schranke für die erlaubte Temperatur.

Reversibilität (1)

R. Landauer
Irreversibility and heat generation in the computing process
IBM Journal on Research and Development, Band 5, S. 183–191, 1961.

Reversibilität (1)

R. Landauer
Irreversibility and heat generation in the computing process
IBM Journal on Research and Development, Band 5, S. 183–191, 1961.

- Irreversible Berechnungen müssen Energie verbrauchen, und zwar pro «weggeworfenem Bit»:
  - $k_B \cdot T \cdot \ln 2 \approx 3 \ldots 4 \cdot 10^{-21}$ Joule
    (Boltzmann-Konstante $k_B \approx 1,3806504 \cdot 10^{-23}\ J/K$)

Reversibilität (2)

C. Bennett
Logical Reversibility of Computation
IBM Journal on Research and Development, Band 17, S. 525–532, 1973.

- Jede berechenbare Funktion kann von einer reversiblen Turingmaschine berechnet werden.
  - Preis, den man zahlt:
    mehr Zeit und/oder mehr Platzbedarf
- Aber man kennt keine reversiblen Technologien für das Rechnen, die keine Energie verbrauchen.

Randbemerkung: reversible Zellularautomaten

- Definition: Die globale Überführungsfunktion ist bijektiv.
- Satz: Dann ist die inverse Funktion sogar immer automatisch auch wieder ein Zellularautomat.
  (Curtis/Hedlund/Lyndon, 1969, Richardson, 1972)
- diverse interessante Ergebnisse, z. B.
  - Reversibilität für lokale δ ein-dimensionaler ZA entscheidbar (Amoroso/Patt, 1972)
  - Reversibilität für lokale δ höher-dimensionaler ZA unentscheidbar (Kari, 1994)
  - Simulationen von irreversiblen durch reversible ZA

Im folgenden werden wir aber beliebige ZA betrachten.

Überblick

Literatur

Roland Vollmar
Some Remarks about the "Efficiency" of Polyautomata.
International Journal of Theoretical Physics, 21(12):1007–1015, 1982.

P. Sanders, R. Vollmar, Th. Worsch
Cellular Automata: Energy Consumption and Physical Feasibility
Fundamenta Informaticae, 52(1–3):231–246, 2002.

Anpassungen des klassischen ZA-Modells

- Ausgangspunkt: klassische r-dimensionale ZA
- Maß für Abwärme: *Zustandsänderungskomplexität* (siehe Kapitel 3)
- *Eingabe:* «zeilenweise» in einen möglichst kleinen *Hyperwürfel* von Zellen
- wesentliche Forderung:
  Abwärme über Oberfläche abführbar, d.h.:
  Es gibt eine Konstante a, so dass
  in jedem Teilwürfel von Zellen mit Seitenlänge d
  während t Schritten nicht mehr als $a(d^r + t \cdot d^{r-1})$
  Zustandsänderungen passieren.
- Solche ZA mögen *überall kalt* heißen.
- Schreibweise CAce

Korollar

Wenn in einem überall kalten ZA in einem Würfel mit Seitenlänge $d(n)$ insgesamt $s(n)$ Zustandsänderungen passieren, dann ist seine Zeitkomplexität mindestens

$$t(n) \in \Omega\left(\frac{s(n) - d(n)^r}{d(n)^{r-1}}\right)$$

Eine untere Schranke für $s(n)$ in einem Teilwürfel liefert also auch eine untere Schranke für die Zeitkomplexität.

Überblick

Definition

- $R = \mathbb{Z}^r$ bezeichne die Menge aller Zellen
- $\langle w \rangle$ bezeichne das Raum-Zeit-Diagramm für Eingabe w,
  i. e. $\langle w \rangle : R \times \mathbb{N} \to Q$ mit $\langle w \rangle(i, \tau) = \Delta^{\tau}(\mathrm{init}(w)))(i)$
- Einschränkung auf Teilmenge M von Zellen:
  $\langle w|M \rangle : M \times \mathbb{N} \to Q$

Lemma

- Sei $R = M_1 \dot{\cup} M_2$ eine Partitionierung der Zellmenge und K die «Grenze» «zwischen» M_1 and M_2:

$$
\begin{aligned}
K \quad = \quad & \{i \in M_1 \mid \exists n \in N : i + n \in M_2\} \\
& \cup \{i \in M_2 \mid \exists n \in N : i + n \in M_1\}
\end{aligned}
$$

- Es seien w_1 und w_2 zwei Wörter gleicher Länge mit $\langle w_1 | K \rangle = \langle w_2 | K \rangle$.

Dann gilt:

- $\langle w_1 | M_1 \rangle \cup \langle w_2 | M_2 \rangle$ ist das Raum-Zeit-Diagramm für eine Eingabe w.
- Falls $0 \in M_i$ ist, ist $w \in L(C) \iff w_i \in L(C)$

Beweis

elementares Rechnen …

Vereinbarung

- Für eine formale Sprache L sei

$$L^{[r]} := \left\{ w \in L \;\middle|\; (1/3) \cdot |w|^{1/r} \in \mathbb{N}_+ \right\}$$

- Wörter aus $L^{[r]}$ füllen den r-dimensional Eingabewürfel vollständig aus (weil $|w|^{1/r} \in \mathbb{N}$).
- Der Faktor $1/3$ ist speziell auf L_{vv} zugeschnitten.

Definition

- Eine *«knapp-über-log»-Funktion* sei eine nicht-fallende Funktion $f : \mathbb{N}_+ \longrightarrow \mathbb{N}_+$ mit

$$\lim_{n \longrightarrow \infty} \log n / f(n) = 0 \ .$$

- von Interesse: *langsam* wachsende $f(n)$
- *Im folgenden bezeichne $f(n)$ stets eine polylogarithmische Funktion.*

Lemma

Für jeden r-dimensionalen ZA, der L_{vv} erkennt, und jede knapp-über-log-Funktion f gilt:
Der ZA macht in dem Teilwürfel mit Seitenlänge $3n^{1/r}$, der in der Mitte den Eingabewürfel enthält,

$$\Omega(n^{(r+1)/r}/f(n))$$

Zustandsänderungen.

Beweis (1)

- ZA erkenne $L^{[r]}$ in Polynomialzeit $t(n)$ (sonst Behauptung trivial)
- o. B. d. A. von-Neumann-Nachbarschaft mit Radius 1
- o. B. d. A. Eingabelänge $n = m^r$
- für $1 \leq i \leq m/6$ betrachte Partitionierungen $M_{i1} \dot\cup M_{i2}$ und zugehörige Grenzen K_i mit

$$
\begin{aligned}
M_{i1} = \{(x_1, \ldots, x_{r-1}, x_r) \mid & -2i \leq x_1 < m + 2i, \\
& \quad \vdots \\
& -2i \leq x_{r-1} < m + 2i \text{ und} \\
& -2i \leq x_r < \frac{m}{3} + 2i\}
\end{aligned}
$$

Beweis (2)

- Alle M_{i1} enthalten das erste Drittel der Eingabe.
- Alle M_{i2} enthalten das dritte Drittel der Eingabe.
- Alle K_i sind paarweise disjunkt mit Größe $\leq b_1 n^{(r-1)/r}$.
- Für $1 \leq i \leq \frac{m}{6}$ und große n ist
die Anzahl der Überquerungsfolgen mit höchstens
$g(n) := \frac{n}{f(n)}$ Zustandsänderungen kleiner oder gleich

$$|Q|^{b_1 n^{(r-1)/r}} \binom{b_1 t(n) n^{(r-1)/r}}{g(n)} |Q|^{g(n)} .$$

Beweis (3)

- Da $f(n)$ knapp-über-log ist, ist $\lim_{n\to\infty} f(n)/n^{1/r} = 0$
- $t(n)n^{(r-1)/r}$ is polynomiell in n
- Also für große n

$$|Q|^{b_1 n^{(r-1)/r}} \binom{b_1 t(n) n^{(r-1)/r}}{g(n)} |Q|^{g(n)}$$

$$\leq \left(b_1 t(n) n^{(r-1)/r}\right)^{g(n)} \left(|Q|^{1+b_1 f(n)/n^{1/r}}\right)^{g(n)}$$

$$\leq \left(2^{\log(b_1 t(n) n^{(r-1)/r})} |Q|^{b_2}\right)^{g(n)} \leq 2^{b_3 g(n) \log n}$$

Beweis (4)

- Es darf nicht Wörter w_1, w_2 und Index i geben mit $\langle w_1 | K_i \rangle = \langle w_2 | K_i \rangle$.
- Für jeden der $m/6$ Werte für i gibt es höchstens $2^{b_3 g(n) \log n}$ Wörter mit höchstens $g(n)$ Zustandsänderungen in K_i
- Für große n gilt

$$\begin{aligned}
(m/6)2^{b_3 g(n) \log n} &= 2^{b_3 g(n) \log n + \log(m/6)} \\
&= 2^{b_3 n (\log n)/f(n) + \log(n^{1/r}/6)} \\
&< 2^{n/3}
\end{aligned}$$

Beweis (5)

- Folglich gibt es für große n mindestens ein Wort, das in allen K_i mehr als $g(n) = n/f(n)$ Zustandsänderungen verursacht.
- dafür Gesamtzahl Zustandsänderungen

$$\frac{mn}{6f(n)} \in \Omega(\frac{n^{(r+1)/r}}{f(n)})$$

Korollar

L_{vv} kann von keinem $\mathbb{Z}^r$-CAce in weniger als $n^{2/r}/f(n)$ Schritten erkannt werden.

Beweis:

- aus $t(n) \in \Omega(\frac{s(n) - d(n)^r}{d(n)^{r-1}})$
- und $s(n) \in \Omega(\frac{n^{(r+1)/r}}{f(n)})$
- folgt $t(n) \in \Omega(\frac{n^{(r+1)/r}}{f(n)n^{(r-1)/r}}) = \Omega(\frac{n^{2/r}}{f(n)})$

Gleich: L_{vv} kann von einem $\mathbb{Z}^r$-CAce in $n^{2/r}$ Schritten erkannt werden.

Lemma

Es sei $d(n)$ raum-konstruierbar in Zeit $t(n)$. Dann gilt:

$$\mathbb{Z}^r\text{-CA-}\textsc{Ext-Time}(d,t) \subseteq \mathbb{Z}^r\text{-CA}\textsc{ce-Ext-Time}(d, d \cdot t)\,.$$

Beweis:

Lemma

Es sei $d(n)$ raum-konstruierbar in Zeit $t(n)$. Dann gilt:

$$\mathbb{Z}^r\text{-CA-}\textsc{Ext-Time}(d, t) \subseteq \mathbb{Z}^r\text{-CAce-}\textsc{Ext-Time}(d, d \cdot t)\,.$$

Beweis: Simuliere die Hyperebenen der Reihe nach

Korollar

Es gibt Probleme die man mit $\mathbb{Z}^r$-CA um einen Faktor $\Omega(\frac{n^{1/r}}{f(n)})$ schneller lösen kann als mit $\mathbb{Z}^r$-CAce.

Beweis: L_{vv}

Satz

Es gibt Probleme, für die $\mathbb{Z}^r$-CAce schneller sind als $\mathbb{Z}^{r-1}$-CA um einen Faktor von mindestens $\Theta(n^{1/(r(r-1))})$.

- Größer kann der Faktor nicht sein, solange die gesamte Berechnung im Eingabewürfel stattfindet.

Beweis

- L_{parity}

Satz

Es gibt Probleme, für die $\mathbb{Z}^{r-1}$-CA im Vergleich zu $\mathbb{Z}^r$-CAce um einen Faktor von mindestens $\Theta(\frac{n^{(r-2)/(r(r-1))}}{f(n)})$ schneller sein können.

Beweis

- L_{vv}

Zusammenfassung

- Physikalisch realisierbare Schaltkreisfamilien (im Sinne von Schorr) können nur polynomiell viel schneller sein als Turingmaschinen.
- «Energieverbrauch» bei Zellularautomaten
  - Zustandsänderungen als Maß dafür
  - Kühlbarkeit ist für manche Probleme eine stärkere Einschränkung als weniger Dimensionen (für andere Probleme nicht)