Schaltkreis-Komplexität

Größe und Tiefe

(a) Sei S ein Schaltkreis.

1. die **Tiefe** von S ist die Länge des längsten Weges im Graph von S,
2. die **Größe** von S ist die Anzahl der Gatter, und
3. der Fanin von S ist das Maximum, über alle Knoten v, der Anzahl eingehender Kanten von v.

(b) Für eine Boolesche Funktion $f : \{0,1\}^n \to \{0,1\}$ ist

$$\mathrm{DEPTH}(f) = \min\{\ d \mid \text{Es gibt einen Schaltkreis der Tiefe } d \text{ für f}\ \}$$

die minimale von f benötigte Tiefe und

$$\mathrm{SIZE}(f) = \min\{\ s \mid \text{Es gibt einen Schaltkreis der Größe } s \text{ für f}\ \},$$

die minimale benötigte Größe, wobei wir jeweils nur Schaltkreise mit den Gattern $\wedge, \vee, \neg$ vom Fanin zwei betrachten.

Ein Abzählargument (1/2)

- Wir bestimmen die Anzahl der Schaltkreise der Größe höchstens s
- und vergleichen dies mit der Anzahl 2^{2^n} der Booleschen Funktionen.

- Wir beschreiben einen Schaltkreis mit m Gattern, indem wir
  - seinen Graphen spezifizieren: Weise jedem Knoten seinen einen oder seine beiden Vorgänger zu und wir erhalten höchstens $m^{2 \cdot m}$ Graphen.
  - die Funktionsweise der Gatter definieren: Bei drei Gattertypen gibt es höchstens 3^m mögliche Funktionsweisen für alle Knoten des Graphen.
  - jeder Quelle eine Eingabeposition zuweisen: Höchstens $n!$ Möglichkeiten.
- Die Anzahl der Schaltkreise mit **genau m Gattern** ist höchstens

$$3^m \cdot m^{2 \cdot m} \cdot n! \leqslant 3^m \cdot m^{2 \cdot m} \cdot m^m \leqslant (3m)^{3m}.$$

Die Anzahl der Schaltkreise mit **höchstens s Gattern** ist höchstens

$$\sum_{m=1}^{s} (3m)^{3m} \leqslant s \cdot (3s)^{3s} = (3s)^{4s}.$$

Ein Abzählargument (2/2)

Es gibt höchstens $(3s)^{4s}$ Schaltkreise mit maximal s Gattern.

Unterschiedliche Funktionen benötigen unterschiedliche Schaltkreise!

- Welche Größe ist notwendig, um mindestens die Hälfte aller Booleschen Funktionen implementieren zu können? Wir müssen fordern

$$2^{2^n-1} \overset{!}{\leqslant} (3s)^{4s},$$

und $s = \Omega(\frac{2^n}{n})$ folgt nach Logarithmieren.

- Schaltkreise der Tiefe t haben höchstens $\sum_{i=0}^{t} 2^i = 2^{t+1} - 1$ Gatter $\Rightarrow$

$$t = n - O(\log n).$$

Für mehr als die Hälfte aller Funktionen $g : \{0,1\}^n \to \{0,1\}$ ist

$$\mathrm{SIZE}(g) = \Omega(\frac{2^n}{n}) \textbf{ und } \mathrm{DEPTH}(g) \geqslant n - O(\log_2 n).$$

Der Status Quo

- Die Funktion f hänge von jeder Eingabe ab:
  - für jedes $i \in \{1, \ldots, n\}$ gibt es eine Eingabe $x \in \{0,1\}^n$, so dass $f(x) \neq f(x \oplus e_i)$, wobei der Vektor e_i in Position i mit 1 übereinstimmt und in allen anderen Positionen nur aus Nullen besteht,

  dann ist $\mathrm{DEPTH}(f) = \Omega(\log_2 n)$ und $\mathrm{SIZE}(f) = \Omega(n)$.
  - Asymptotisch bessere untere Schranken sind nicht bekannt :-((

- Für jede Boolesche Funktion $f : \{0,1\}^n \to \{0,1\}$:
  - $\mathrm{DEPTH}(f) \leqslant n + \lceil \log_2 n \rceil$,
    - ⋆ Entweder hat f oder $\neg f$ eine DNF mit höchstens 2^n Monomen.
    - ⋆ Eine DNF für f benötigt Tiefe höchstens $\lceil \log_2 n \rceil$ für die Monome und Tiefe maximal n für die Disjunktion der höchstens 2^n Monome.
  - $\mathrm{SIZE}(f) = O(2^n/n)$: Übungsaufgabe.

Untere und obere Schranken für Tiefe und Größe klaffen extrem weit auseinander :-((

$\mathsf{P}_{/\mathrm{poly}}$

Schaltkreise polynomieller Größe

$$\mathrm{P}_{/\mathrm{poly}} = \bigcup_{k \in \mathbb{N}} \mathrm{SIZE}(n^k)$$

ist die Klasse aller Sprachen mit (nicht-uniformen) Schaltkreisen polynomieller Größe.

(a) Für alle Funktionen $t : \mathbb{N} \to \mathbb{N}$ ist

$$\mathrm{DTIME}(t) \subseteq \mathrm{SIZE}(t^2).$$

Das haben wir implizit im P-Vollständigkeitsergebnis für CVP festgestellt.

(b) $\mathrm{P} \subseteq \mathrm{P}_{/\mathrm{poly}}$.

(c) Wenn $L \notin \mathrm{P}_{/\mathrm{poly}}$ für eine Sprache $L \in \mathrm{NP}$, dann gilt $\mathrm{P} \neq \mathrm{NP}$.

Die Herleitung super-polynomieller Größenschranken für ein Problem in NP ist sehr schwer.

Was tun?

Die Trennung von P und NP ist eine zentrale Fragestellung der theoretischen Informatik.

Wir „üben“, indem wir super-polynomielle Größenschranken zeigen für

- Schaltkreise mit beschränkter Tiefe, aber unbeschränktem Fanin
- und monotone Schaltkreise.

Schaltkreise beschränkter Tiefe

AC^0

AC^0 ist die Klasse aller Sprachen mit uniformen Schaltkreisen polynomieller Größe und konstanter Tiefe, aber unbeschränktem Fanin.

Zu AC^0 gehören

- die Addition von zwei n-bit Zahlen,
- die Multiplikation Boolescher Matrizen,
- für jedes $m \in \mathbb{N}$ und jedes $k \leqslant \log_2^m n$ die Fragen, ob ein String von n Nullen und Einsen mindestens, höchstens oder genau k Einsen besitzt.

Nicht zu AC^0 gehören

- die Paritätsfunktion $\texttt{xor}_n(x) = x_1 \oplus \cdots \oplus x_n$,
- die Mehrheitsfunktion $\texttt{majority}_n(x) = 1 \Leftrightarrow \sum_{i=1}^{n} x_i \geqslant n/2$,
- die Multiplikation von zwei n-bit Zahlen,
- die Frage, ob ein ungerichteter regulärer Graph vom Grad zwei zusammenhängend ist.

Sind Funktionen in AC^0 strukturell eingeschränkt?

Sensitivity (1/2)

$e_i \in \{0,1\}^n$ ist das Wort mit einer Eins und zwar in Position i.

$f : \{0,1\}^n \to \{0,1\}$ ist eine Boolesche Funktion.

- Die Empfindlichkeit e_x von f für Eingabe x ist die Anzahl der Bitpositionen i mit $f(x) \neq f(x \oplus e_i)$.
- Die **Empfindlichkeit** von f ist die durchschnittliche Empfindlichkeit e_x,

$$e = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} e_x.$$

- Wir werden zeigen, dass alle Funktionen in AC^0 eine höchstens poly-logarithmische Empfindlichkeit besitzen.

- Die Paritätsfunktion $\mathtt{xor}_n(x) = \oplus_{i=1}^n x_i$ hat maximale Empfindlichkeit n und gehört deshalb nicht zu AC^0.
- Die Mehrheitsfunktion und die Multiplikation besitzen eine viel zu hohe Empfindlichkeit und gehören nicht zu AC^0.

Sensitivity (2/2)

Die zentrale Frage:

Warum berechnen „AC^0-Schaltkreise“,

also Schaltkreise beschränkter Tiefe und polynomieller Größe

nur Funktionen mit höchstens poly-logarithmischer Empfindlichkeit?

Sei S ein AC^0-Schaltkreis.

- Wir führen zuerst eine Schönheitsoperation auf S aus
- und führen dann die Methode der **Restriktionen** ein, um die Schwächen von S offenzulegen.

Eine Schönheitsoperation

Wir können annehmen, dass

- Negationsgatter nur für Eingabegatter verwandt werden und dass
- der Schaltkreis in Schichten aufgeteilt ist, wobei
  - Kanten nur zwischen benachbarten Schichten verlaufen,
  - eine Schicht entweder eine UND-Schicht ist (alle Gatter einer Schicht sind UND-Gatter) bzw. eine ODER-Schicht (alle Gatter sind ODER-Gatter)
  - und dass der Schaltkreis alterniert: Eine UND-Schicht folgt auf eine ODER-Schicht und umgekehrt.

Übungsaufgabe:
Zeige, dass jeder Schaltkreis in eine solche Form gebracht werden kann, wobei die Größe höchstens quadriert wird und die Tiefe nicht ansteigt.

Restriktionen

Restriktionen

Eine Restriktion ist eine Funktion

$$\rho : \{1, \ldots, n\} \to \{0, 1, *\}.$$

- Wir wenden ρ auf den Schaltkreis S an und erhalten den Schaltkreis

$$S|_\rho.$$

- $S|_\rho$ erhält nicht die Eingabe $x \in \{0,1\}^n$, sondern die Eingabe $\rho(x)$.
  - Wenn $\rho(i) = 0$ oder $\rho(i) = 1$, dann erhält S nur den auf 0, bzw 1 eingefrorenen Wert von x_i.
  - Wenn $\rho(i) = *$, dann erhält S den unveränderten Wert x_i.
    Wir sagen, dass Position i „gesternt“ wird.

Unsere Strategie

Wir nehmen an, dass die unterste Schicht von Schaltkreis S aus UND-Gattern besteht: Jedes Gatter der Tiefe zwei berechnet also eine DNF.

- Wir würfeln eine zufällige Restriktion ρ aus, um **jede** DNF der zweiten Schicht in eine nicht zu große KNF zu überführen.
  - Warum wählen wir Restriktionen zufällig aus?
  - Für eine einzige DNF können wir Restriktionen maßschneidern, aber wie sollen wir das bei vielen DNFs tun?
- Danach kann die zweite und dritte Schicht verschmolzen werden: Die Tiefe unseres Schaltkreises ist um Eins gesunken!

- Wir wiederholen unser Vorgehen und müssen am Ende einen Schaltkreis der Tiefe zwei analysieren.

Nicht so schnell!

- Legen Restriktionen denn Schwächen des Schaltkreises offen?
- Explodiert denn nicht die Größe, wenn wir eine DNF in eine KNF umwandeln? Warum können denn Restriktionen diese Explosion verhindern?

Tiefe zwei: Von DNFs zu KNFs (1/2)

Angenommen, der Schaltkreis S ist eine DNF mit Ausgabegatter g.

- Wenn der Fanin der UND-KINDER von g **anfänglich klein** war:
  - Einige UND-Kinder überleben, aber bei vielen UND-Kindern sollte mindestens eines der UND-Kinder nur Einsen erhalten:

    Wir haben g trivialisiert und können auch g ersatzlos streichen.

- Wenn der Fanin der UND-Kinder **anfänglich nicht klein** war: **Nicht-entschiedene** UND-Kinder, also UND-Kinder, die nur Einsen *und* Sterne erhalten, machen uns das Leben schwer.
  - Unsere Hoffnung: Die Anzahl dieser Kinder ist klein und in diesem Fall hängt der Eltern-Knoten g von hoffentlich nur wenigen Sternen ab.
  - Dann könnte die DNF von g in eine nicht zu große KNF überführt werden. und S hätte eine Schicht weniger:

    die Schicht von g wäre zu einer UND-Schicht „gemorpht“ und könnte mit der UND-Schicht der Eltern von g verschmolzen werden.

Tiefe zwei: Von DNFs zu KNFs (2/2)

Die Überführung einer DNF g in eine äquivalente KNF g' kann teuer werden.

Betrachte die simple DNF

$$g = \bigvee_{i=1}^{n/2} x_{2i-1} \wedge x_{2i}.$$

- Die KNF g' muss
  - alle $2^{n/2}$ Eingaben, die jedes Interval $\{2i-1, 2i\}$ mit genau einer Eins treffen, ausgeschalten und
  - alle Eingaben, die irgendein Intervall mit zwei Einsen treffen, akzeptieren.
    - Jede Klausel von g' muss mindestens ein positives Literal x_j für $j \in \{2i-1, 2i\}$ besitzen.
    - Nur eine einzige Eingabe, die jedes Intervall mit genau einer Eins trifft, kann ausgeschaltet werden.

g' hat mindestens $2^{n/2}$ Klauseln!

Kritische Sterne

Die Überführung einer DNF $g|_\rho$ in eine KNF g' kann exponentielle Größe in der Anzahl „kritischer Sterne“ verschlingen.

- Nenne einen Stern **kritisch**, wenn die von $g|_\rho$ berechnete Funktion vom Wert der gesternten Variable abhängt.
- Hoffentlich gibt es nach Anwendung der Restriktion nur wenige kritische Sterne!

Aufgabe: Wie bestimmt man die Anzahl kritischer Sterne?

Der kanonische Entscheidungsbaum und die Anzahl kritischer Sterne

Der kanonische Entscheidungsbaum T_ρ (1/2)

ρ sei eine Restriktion, die wir auf die DNF g anwenden.

- Der Baum T_ρ besteht nur aus der Wurzel, wenn
  - $g|_\rho$ entweder konstant Null (alle Monome wurden eliminiert) oder
  - konstant Eins (ein Monom wurde erfüllt) ist.

  Beschrifte die Wurzel mit dem Wert von $g|_\rho$.
- Und wenn die Restriktion ρ die DNF nicht trivialisiert?
  - Ein Monom M_i ist entweder **tot** (M_i wurde durch ρ falsifiziert) oder **lebendig** (die Literale von M_i erhalten keine Null, aber mindestens einen Stern).
  - Lege eine beliebige Reihenfolge auf den Monomen und Variablen fest.

Mit $T\rho$ möchten wir messen, von wievielen Variablen $g|_\rho$ abhängt.

Der kanonische Entscheidungsbaum T_ρ (2/2)

Sei M_i das erste lebendige Monom.

- Durchlaufe die gesternten Variablen von M_i gemäß der festgelegten Reihenfolge.
- Ist $l = (\neg)x_j$ in M_i gesternt, dann verzweigt T_ρ nach dem Wert von l.
  - Wenn $l = 1$
    - fahre rekursiv mit der nächsten gesternten Variable von M_i fort.
    - Gibt es keine solche Variable, endet unsere Konstruktion, und wir markieren das Blatt mit dem Wert 1: $g|_\rho$ wurde erfüllt.
  - Wenn $l = 0$
    - fahre rekursiv mit dem nächsten lebendigen Monom M_j fort, denn wir haben gerade das Monom M_i falsifiziert.
    - Gibt es kein solches Monom M_j, endet unsere Konstruktion, und wir markieren das Blatt mit dem Wert Null: Wir haben alle Monome von $g|_\rho$ falsifiziert.

Die Tiefe von T_ρ

Angenommen, der kanonische Entscheidungsbaum T_ρ hat die Tiefe s.

- T_ρ ist zu einer KNF mit Monomen der Länge höchstens s äquivalent.
  - Betrachte einen Weg W in T, der in einem 0-Blatt endet.
  - Formuliere eine Klausel K_W der Länge s, die wenn erfüllt sicherstellt, dass W nicht durchlaufen wird.
  - Die Konjunktionen aller Klauseln K_W für sämtliche 0-Wege W ist die von uns gewünschte KNF g' mit Klauseln der Länge höchstens s.
- T_ρ ist zu einer DNF mit Monomen der Länge höchstens s äquivalent.
  - Für jeden 1-Weg W bilden wir das Monom M_W, das nur erfüllt wird, wenn W durchlaufen wird.
  - Die gewünschte DNF ist die Disjunktion aller Monome M_W für 1-Wege W.

Ein Entscheidungsbaum der Tiefe s ist also
zu einer DNF wie auch zu einer KNF mit Bottom-Fanin s äquivalent.

Das Switching Lemma

Der Bottom-Fanin

Wir fordern, dass der „Bottom-Fanin“ t,

also der Fanin der untersten Schicht des Schaltkreises S

moderat groß ist.

- Und wenn der Bottom-Fanin von S anfänglich größer als t ist?
- Führe eine zusätzliche Schicht ein und erhöhe die Tiefe um höchstens Eins.

Die wichtigen Parameter

- **n** ist die Anzahl der Variablen,
- **l** ist die Anzahl der Sterne,
- **t** ist der maximale Bottom Fanin.
- **s** ist die Tiefe von T_ρ.

Wann schneiden sich Mengen hochwahrscheinlich?

- Die Wahrscheinlichkeit, dass eine zufällig ausgewürfelte Menge der Größe l eine vorgegebene Menge der Größe t verfehlt, ist

$$p = \binom{n-t}{l} / \binom{n}{l}.$$

  - Es ist $p = \frac{n-t}{n} \cdots \frac{n-t-l+1}{n-l+1} \approx (\frac{n-t}{n})^l = (1-t/n)^l \approx e^{-tl/n}$.
  - Aber $e^{x/(1-x)} \leqslant 1 - x \leqslant e^{-x}$ gilt.

- Falls $lt << n$, ist

$$\mathbf{p \approx 1 - \frac{lt}{n}},$$

$\Rightarrow$ $\mathbf{\frac{lt}{n}}$ ist ungefähr die Wahrscheinlichkeit, dass ein vorgegebenes Monom von mindestens einem Stern getroffen wird.

Das Switching Lemma

- S ist ein Schaltkreis der Tiefe zwei mit Bottom-Fanin höchstens t.
- Eine Restriktion ρ wird zufällig ausgewählt,
  - indem $l \leqslant n/3$ Positionen für die Platzierung der Sterne ausgewürfelt werden
  - und in den restlichen Positionen unabhängig voneinander eine Null oder Eins jeweils mit Wahrscheinlichkeit $1/2$ zugewiesen wird.

Dann gilt

$$\mathrm{pr}[\,\mathrm{depth}(S|_\rho) \geqslant s\,] \leqslant (\frac{12lt}{n})^s.$$

Wenn die Wahrscheinlichkeit l/n für einen Stern so niedrig ist, dass die meisten Monome verfehlt werden, dann ist mit hoher Wahrscheinlichkeit entweder ein Monom erfüllt oder es gibt nur wenige lebendige Monome.

Switching Lemma, der Beweis (1/5)

- O.B.d.A. ist S eine DNF.
- Der kanonische Entscheidungsbaum T_ρ habe die Tiefe mindestens **s**.
- Den lexikografisch ersten, in der Wurzel von T_ρ beginnenden Weg der Länge **s** nennen wir π.

- Wir ändern die Restriktion ρ ab, und nennen die neue Restriktion ρ'.
  - Die in ρ (auf Null oder Eins) gesetzten Variablen werden in ρ' nicht verändert.
  - **s** in ρ gesternte Variablen werden in ρ' auf Null oder Eins gesetzt.
    - ★ Welche gesternten Variablen werden gesetzt?
    - ★ Das wird uns der Weg π erzählen.

- Und was ist der Clou?
  - Bei kleiner Wahrscheinlichkeit t/n für einen Stern ist die Anzahl der abgeänderten Restriktionen ρ' viel kleiner als die Anzahl der Restriktionen ρ.

    Warum ist das so?

Switching Lemma, der Beweis (2/5)

Gibt es wirklich weniger Restriktionen ρ': Zwar hat ρ' genau s Sterne weniger, aber s zusätzliche Variablen müssen auf Null oder Eins gesetzt werden!

- Wieviele verschiedene Restriktionen gibt es?
  - Genau $\binom{n}{l}2^{n-l}$ Restriktionen, gegenüber $\binom{n}{l-s}2^{n-(l-s)}$ Restriktionen ρ'.
  - Die Ersparnis ist für $l = o(n)$ signifikant, denn

$$\frac{\binom{n}{l-s}2^{n-(l+s)}}{\binom{n}{l}2^{n-l}} \leqslant \left(\frac{l}{n-l}\right)^s 2^s = \left(\frac{2l}{n-l}\right)^s.$$

- Aber wir können nicht einfach Sterne in ρ' willkürlich setzen:
  - Wir müssen die schlechten Restriktionen ρ zählen,
  - d.h. wir müssen ρ aus ρ' rekonstruieren können.

Wird ρ mit **geringer** Zusatzinformation aus ρ' rekonstruiert, dann

- gibt es nur wenige schlechte Restriktionen ρ, also Restriktionen mit großer Tiefe.
- $\Rightarrow$ Eine große Tiefe **s** ist unwahrscheinlich.

Switching Lemma, der Beweis (3/5)

Wir ändern ρ in einer Reihe von Schritten.

- Wir beginnen mit dem „ersten“ lebendigen Monom M_1 für ρ und **setzen die Werte der von ρ gesternten Variablen so, dass M_1 erfüllt wird**.
  - M_1 ist das **erste erfüllte** Monom und bleibt das erste erfüllte Monom auch wenn später weitere Sterne von ρ gesetzt werden.
  - M_1 ist somit als erstes durch ρ' erfülltes Monom eindeutig bestimmt!
  - Wenn M_1 nur wenige Literale besitzt, dann kann die Position eines Sterns in M_1 mit **geringer Zusatzinformation** beschrieben werden.
- Wir übernehmen diese Änderungsstrategie für alle weiteren, zum Zeitpunkt ihrer Behandlung lebendigen Monome M auf dem Weg π.

Switching Lemma, der Beweis (4/5)

Welche Zusatzinformation wird benötigt?

1. Wir sind möglicherweise vom Weg π abgewichen. Wie finden wir zurück?
   - Wenn M_1 genau s_1 Sterne besitzt, dann gibt es genau 2^{s_1} Möglichkeiten für die von π gesetzten Variablen: Eine dieser 2^{s_1} Möglichkeiten müssen wir angeben.
   - Wenn das $j$te Monom M_j zum Zeitpunkt seiner Behandlung s_j Sterne besitzt, dann beschreibt der Vektor

$$\text{Zurück} \in \{0,1\}^{s_1+s_2\cdots} = \{0,1\}^{s}$$

   wie wir für jedes M_j zu π zurückfinden. Zurück ist einer von 2^s Vektoren.
2. Wir geben für alle M_j an, welche ihrer Variablen in ρ zum Zeitpunkt der Behandlung von M_j gesternt waren.

Wir arbeiten mit kurzen Monomen, und können deshalb die gesetzten Sterne billig rekonstruieren.

Switching Lemma, der Beweis (5/5)

Wieviele Möglichkeiten gibt es für die Platzierung der Sterne in lebendigen Monomen?

$$\text{stars}(t, s) = \{ (\beta_1, \ldots, \beta_k) \quad | \quad \forall i : \beta_i \in \{*, -\}^t, \text{ jedes } \beta_i \text{ besitzt mindestens 1 Stern bei insgesamt } s \text{ Sternen}\}.$$

- β_j beschreibt die Platzierung der Sterne in M_j.
- Ein Vektor „Sterne“ $\in \text{stars}(t, s)$ beschreibt die in ρ' gesetzten Sterne.
  $|\text{stars}(t, s)|$ ist die Anzahl der Möglichkeiten für die Wahl der Sterne in den lebendigen Monomen.

Die Zuordnung

$$\rho \mapsto (\rho', \text{Zurück}, \text{Sterne})$$

ist injektiv: ρ ist aus ρ', „Zurück“ und „Sterne“ rekonstruierbar.

$|\text{stars}(t,s)| < (t/\ln(2))^s$.

Die reelle Zahl γ erfülle die Gleichung $(1+1/\gamma)^t = 2$. Wir zeigen

$$|\text{stars}(t,s)| \leqslant \gamma^s$$

durch Induktion über s.

- Die Basis: $|\text{stars}(t,0)| = 0$ und die Ungleichung ist für $s = 0$ richtig.
- Der Induktionsschritt:
  - Wenn β_1 genau i Sterne besitzt, dann hat $(\beta_2, \ldots, \beta_k)$ genau $s - i$ Sterne.
  - Es gibt höchstens $\binom{t}{i}$ Platzierungen der i Sterne in β_1.
  - Mit der Induktionsannahme ist

$$\begin{aligned} |\text{stars}(t,s)| &= \sum_{i=1}^{\min\{t,s\}} |\text{stars}(t,s-i)| \leqslant \sum_{i=1}^{t} \binom{t}{i} \gamma^{s-i} \\ &= \gamma^s \sum_{i=1}^{t} \binom{t}{i} (\frac{1}{\gamma})^i = \gamma^s \left((1+\frac{1}{\gamma})^t - 1 \right) = \gamma^s. \end{aligned}$$

- Es ist $(1+1/\gamma)^t < e^{t/\gamma} \Rightarrow 2 < e^{t/\gamma}$. Nach Logarithmierung folgt $\gamma < t/\ln(2)$.

Wieviele schlechte Restriktionen ρ gibt es?

(also Restriktionen ρ für die der Entscheidungsbaum T_ρ eine Tiefe von mindestens s besitzt)

Wir beschreiben ρ durch ρ', den Vektor Zurück $\in \{0,1\}^s$ und den Vektor „Sterne“ (als einen von höchstens $(\frac{t}{\ln(2)})^s$ Vektoren). $\Rightarrow$ Es gibt höchstens

$$\binom{n}{l-s} 2^{n-(l-s)} \cdot 2^s \cdot (\frac{t}{\ln(2)})^s$$

schlechte Restriktionen, aber $\binom{n}{l} 2^{n-l}$ Restriktionen mit l Sternen.
Die Wahrscheinlichkeit einer schlechten Restriktion ρ ist für $l \leqslant n/3$

$$\leqslant \frac{\binom{n}{l-s} 2^{n-(l-s)} \cdot 2^s \cdot (\frac{t}{\ln(2)})^s}{\binom{n}{l} 2^{n-l}} = \frac{\binom{n}{l-s} 2^{n-l} \cdot (\frac{4t}{\ln(2)})^s}{\binom{n}{l} 2^{n-l}} = \frac{\binom{n}{l-s} (\frac{4t}{\ln(2)})^s}{\binom{n}{l}}$$

$$\leqslant (\frac{l}{n-l})^s \cdot (\frac{4t}{\ln(2)})^s \leqslant (\frac{8lt}{n\ln(2)})^s \leqslant (\frac{12lt}{n})^s.$$

Wir machen den Schaltkreis platt (1/2)

Der Schaltkreis S habe die Tiefe d und Größe g. (Wir erhöhen die Tiefe um Eins und können annehmen, dass der Bottom-Fanin höchstens $t = 2\log_2 g$ beträgt.)

- Reduziere die Tiefe von S um Eins: Wende das Switching Lemma an mit Sternwahrscheinlichkeit
$$p = \frac{l}{n} = \frac{1}{24t}$$
  - Die Wahrscheinlichkeit, dass eines der höchstens g vielen DNFs einen Entscheidungsbaum der Tiefe mindestens $s = 2\log_2 g$ besitzt, ist höchstens
$$g\left(\frac{12lt}{n}\right)^{2\log_2 g} = g\left(\frac{12}{24}\right)^{2\log_2 g} = \frac{g}{g^2} = \frac{1}{g}.$$
  - Wir wandeln mit Wahrscheinlichkeit $1 - \frac{1}{g}$ **alle** DNFs in KNFs der Länge $\leqslant t$ um. Es verbleiben $l = \frac{n}{24t}$ gesternte, also ungesetzte Variablen.
- Wiederhole $d - 2$ Mal: Die Tiefe ist höchstens zwei mit Bottom-Fanin $2\log_2 g$. Es verbleiben $n/(24t)^{d-2}$ ungesetzte Variablen.

Wir machen den Schaltkreis platt (2/2)

Sei S ein Schaltkreis der Tiefe d und Größe g.

- S ist nach Anwendung zufälliger Restriktionen mit Wahrscheinlichkeit mindestens $1 - \frac{d}{g}$ äquivalent zu einem Schaltkreis der Tiefe zwei mit Bottom-Fanin höchstens $2 \log_2 g$.
- Genau $n/(48 \log_2 g)^{d-2}$ Variablen sind ungesetzt.

- Wir haben den Schaltkreis platt gemacht.
- Abhängig von der Größe g und der Tiefe d von S bleiben noch viele Variablen ungesetzt.

Viele Funktionen besitzen aber nur dann eine eine DNF oder KNF, wenn der Bottom-Fanin (fast) so groß ist wie die Anzahl freier Variablen.

Ja, und?

Eine untere Schranke für die Paritätsfunktion

Die Paritätsfunktion $\mathtt{xor}_n(x) = x_1 \oplus \cdots \oplus x_n$

- Wenn wir eine Restriktion mit m Sternen auf $\mathtt{xor}_n$ anwenden, erhalten wir mit $\mathtt{xor}_m$ oder $1 \oplus \mathtt{xor}_m$ eine weitere Paritätsfunktion.
- Eine DNF oder KNF erkennt eine Paritätsfunktion nur dann fehlerfrei, wenn der Bottom-Fanin mit der Anzahl freier Variablen übereinstimmt.
  (Warum? Ein Monom, dessen Literalzahl kleiner ist als die Anzahl freier Variablen, akzeptiert Eingaben gerader wie auch ungerader Parität, der Schaltkreis arbeitet also inkorrekt.)
- Deshalb folgt

$$\begin{aligned} 2\log_2 g &\geqslant n/(48\log_2 g)^{d-2} \text{ bzw,} \\ n &\leqslant (48\log_2 g)^{d-1}. \end{aligned}$$

Jeder Schaltkreis der Tiefe $d \geqslant 2$ für die Paritätsfunktion $\mathtt{xor}_n$ benötigt mindestens

$$2^{\Omega(n^{1/(d-1)})}$$

Gatter. Übungsaufgabe: Größe $2^{O(n^{1/(d-1)})}$ reicht aus.

Sensitivity

Sensitivity: Schaltkreise der Tiefe zwei

Der Schaltkreis S sei eine DNF oder eine KNF mit Bottom-Fanin höchstens t. Dann besitzt S die durchschnittliche Empfindlichkeit $\mathbf{e \leqslant 2t}$. Beachte, dass wir keine Aussage über die Größe der DNF oder KNF machen, die Fanin-Schranke t allein ist schon ausreichend.

O.B.d.A. gelte

$$S = \bigvee_{\vec{i}} \bigwedge_{j=1}^{t} (\neg) x_{i_j}.$$

Die Klammern drücken aus, dass eine Variable „möglicherweise“ negiert auftritt.

- Wenn S die Eingabe x akzeptiert, dann erfüllt x ein Monom $\bigwedge_{j=1}^{t} (\neg) x_{i_j}$.
- Aber alle „Nachbarn“ $x \oplus e_k$ für $k \notin \{i_1, \ldots, i_t\}$ werden auch akzeptiert und x besitzt die Empfindlichkeit $e_x \leqslant t$.
  - Wir stellen uns die Eingaben als Knoten des n-dimensionalen Würfels vor.
  - Sei E die Menge der Kanten, die eine 1-Eingabe mit einer 0-Eingabe verbinden und Eins die Menge der 1-Eingaben. Dann gilt

$$e = 2|E|/2^n \leqslant 2t \cdot \mathrm{Eins}/2^n \leqslant 2t$$

für die Empfindlichkeit e von S.

Sensitivity: Schaltkreise beliebiger Tiefe (1/2)

Zur Erinnerung:

Ein Schaltkreis S der Tiefe d und Größe g kann mit Wahrscheinlichkeit $\geqslant 1 - \frac{d}{g}$ in einen Schaltkreis S' der Tiefe zwei mit Bottom-Fanin $2\log_2 g$ und genau $n/(48\log_2 g)^{d-2}$ ungesetzten Variablen überführt werden.

- S' besitzt die durchschnittliche Empfindlichkeit höchstens $\mathbf{4\log_2 g}$.
- Wenn S also die Empfindlichkeit e besitzt, dann wird man erwarten, dass e nach Anwendung der Restriktion(en) auf

$$\approx \frac{[\mathbf{e}] \cdot [\mathbf{n}/(\mathbf{48}\log_2 \mathbf{g})^{\mathbf{d-2}}]}{\mathbf{n}} = \frac{\mathbf{e}}{(\mathbf{48}\log_2 \mathbf{g})^{\mathbf{d-2}}}$$

fällt.

  - Wenn wir eine Teilmenge $B \subseteq \{1, \ldots, n\}$ der Mächtigkeit b gemäß der Gleichverteilung zufällig auswürfeln, dann ist

$$E[\ |A \cap B|\ ] = \Theta(\frac{a \cdot b}{n})$$

    die erwartete Größe von $|A \cap B|$ für eine fest gewählte Menge A mit $|A| = a$.

Sensitivity: Schaltkreise beliebiger Tiefe (2/2)

S' besitzt eine Empfindlichkeit von höchstens $4 \log_2 g$ und erbt die Empfindlichkeit $\approx \frac{e}{(48 \log_2 g)^{d-2}}$ von S.

Wir erhalten die Bedingung

$$4 \log_2 g = \Omega\left(\frac{e}{(48 \log_2 g)^{d-2}}\right).$$

Ein Schaltkreis der Tiefe d und Größe g berechnet eine Funktion mit durchschnittlicher Empfindlichkeit höchstens

$$\mathbf{O}(\log_2 \mathbf{g})^{\mathbf{d}-\mathbf{1}}.$$

Schaltkreise konstanter Tiefe und polynomieller Größe besitzen also höchstens die durchschnittliche Empfindlichkeit $\log_2^{\mathbf{O(1)}} \mathbf{n}$.

Schaltkreise und Würfelzerlegungen

Restriktionen und Würfelzerlegungen

Unser Ziel ist eine Zerlegung des n-dimensionalen Würfels in „wenige“ Teilwürfel, so dass der Schaltkreis S auf jedem Teilwürfel konstant ist.

$\texttt{xor}_n$ *ist nur auf ein-elementigen Teilwürfeln konstant. Wir erhalten eine Aussage darüber wie schlecht* $\texttt{xor}_n$ *durch einen zu kleinen Schaltkreis S* ***approximiert*** *wird.*

- Teilwürfel und Restriktionen entsprechen sich ein-eindeutig: Wir können den Schaltkreis S über einem Teilwürfel auswerten oder eine entsprechende Restriktion auf S anwenden.
  - Für eine DNF produziert jeder Entscheidungsbaum mit b Blättern eine Zerlegung in b Teilwürfel, so dass die DNF auf jedem Teilwürfel konstant ist.
- Wir produzieren eine Würfelzerlegung mit Hilfe von Restriktionen,
  - indem wir zuerst Restriktionen ρ zufällig auswürfeln
  - und dann einen Entscheidungsbaum T_ρ für **alle** DNF's von S bauen.

Welche Eigenschaften soll T_ρ haben?

Der gemeinsame Entscheidungsbaum T_ρ (1/4)

- Der Schaltkreis S besitze die DNFs $S_i|_\rho$ in seiner zweiten Schicht.
- Der Bottom-Fanin von S ist t.

Das Ziel. Für jeden in der Wurzel von T_ρ beginnenden Weg W:

- W wird beschrieben durch eine Restriktion ρ_W und
  KNFs K_W mit Bottom-Fanin t und damit durch **eine** KNF mit Bottom-Fanin t.
- Auf W ist nach Anwendung der Restriktion ρ_W und der KNF K_W **jede** DNF $S_i|_\rho$ zu einer KNF mit Bottom-Fanin t äquivalent.

  Verschmelze die zweite mit der dritten Schicht von S.

Die zentrale Frage:

Mit welcher Wahrscheinlichkeit (über die Wahl von ρ) ist die Tiefe von T_ρ wie groß?

Der gemeinsame Entscheidungsbaum T_ρ (2/4)

Nach Abarbeiten der ersten $i-1$ DNF's haben wir den Baum T_ρ^{i-1} konstruiert.

- Für ein Blatt v von T_ρ^{i-1} ist $T_\rho^v(S_i)$ der kanonische Entscheidungsbaum von $S_i|_\rho$ (die auf dem Weg nach v gesetzten Variablen beeinflussen die Konstruktion von $T_\rho^v(S_i)$)
- Belegungen verzweigen in v nach der Länge ihres Weges in $T_\rho^v(S_i)$.
  - Belegungen mit Wegen der Länge $\leqslant t$ in $T_\rho^v(S_i)$ „laufen" zum linken Kind v_0.
    - ★ Wir nennen diese Belegungen „kurz" für $S_i|_\rho$.
    - ★ $S_i|_\rho$ ist für kurze Belegungen mit einer KNF vom Bottom-Fanin t äquivalent (ein Entscheidungsbaum der Tiefe $\leqslant t$ beschreibt die Menge kurzer Belegungen von v_0).
    - ★ Die Konstruktion von T_ρ^i endet für kurze Belegungen in v_0 :-))
  - Belegungen mit Wegen der Länge $> t$ in $T_\rho^v(S_i)$, laufen zum rechten Kind v_1.
    - ★ Wir nennen diese Belegungen „lang" für $S_i|_\rho$.
    - ★ Wir kleben $T_\rho^v(S_i)$ an das Kind v_1 und entfernen alle nur von kurzen Belegungen durchlaufenen Teilbäume.
    - ★ $S_i|_\rho$ ist für lange Belegungen mit einer KNF vom Bottom-Fanin Null äquivalent, aber kann die Tiefe explodieren? :-((

Wie wird in v verzweigt?

Der gemeinsame Entscheidungsbaum T_ρ (3/4)

Wir müssen durch KNFs $k_i^v(x)$ und $l_i^v(x)$ feststellen, ob
eine Belegung x einen kurzen oder einen langen Weg in $S_i|_\rho$ beschreibt.

- Wie sieht k_i^v aus?
  - Die Konjunktion σ_i beschreibe den $i$ten, in der Wurzel von $T_\rho^v(S_i)$ beginnenden Weg der Länge genau t, der nicht in einem Blatt endet. Setze

$$k_i^v = \bigwedge_i \neg\sigma_i.$$

- Wie sieht l_i^v aus?
  - Die Konjunktion μ_i beschreibe den in der Wurzel von $T_\rho^v(S_i)$ beginnenden Weg zum $i$ten Blatt der Tiefe $\leqslant t$. Setze

$$l_i^v = \bigwedge_i \neg\mu_i.$$

Sowohl k_i^v wie auch l_i^v sind KNFs mit einen Bottom-Fanin von höchstens t.

Der gemeinsame Entscheidungsbaum T_ρ (4/4)

Das Ziel ist erreicht:

- Jeder Weg W in T_ρ wird durch eine Restriktion ρ_W und eine KNF vom Bottom-Fanin t beschrieben.
- Auf W ist nach Anwendung der Restriktion ρ_W und der KNF K_W jede DNF $S_i|_\rho$ zu einer KNF mit Bottom-Fanin t äquivalent.

  Wir können die zweite mit der dritten Schicht von S verschmelzen.

Die zentrale Frage bleibt:

Mit welcher Wahrscheinlichkeit (über die Wahl von ρ) ist die Tiefe von T_ρ wie groß?

Ein Switching Lemma für eine ganze Schicht

$(S_i \mid 1 \leqslant i \leqslant g)$ sei eine Folge von DNFs mit Bottom-Fanin $\leqslant t = 2 \log_2 g$.
Für die Anzahl $l = \Theta(\frac{n}{t})$ von Sternen hat T_ρ mit Wahrscheinlichkeit höchstens

$$O(\frac{lt}{n})^s$$

einen Weg der Länge mindestens s.

Warum ist das „neue“ Switching Lemma richtig?

Ein Switching Lemma für eine ganze Schicht

Der Weg W in T_ρ besitze die Länge s.

- Können wir die Beweismethode des Switching Lemmas anwenden?
  - Wir müssen nicht nur die veränderte Restriktion ρ' angeben,
  - beschreiben welche Variablen eines erfüllten Monoms in ρ gesternt waren
  - und festlegen, wie wir zu W zurückfinden,
  - sondern jedes i, so dass W „lang“ für $S_i|_\rho$ ist, muss offengelegt werden.

- Das (konventionelle) Switching Lemma beschränkt die Wahrscheinlichkeit, dass $T_\rho^v(S_i)$ eine Tiefe größer als $2\log_2 g$ besitzt, auf $O(lt/n)^{2\log_2 g}$.

  Der „Profit” $O(lt/n)^{2\log_2 g}$ übertrifft die „Offenlegungs-Kosten“ g, solange $O(lt/n)$ genügend klein ist.

Das neue Switching Lemma ist eine direkte Konsequenz des alten.

Wieviele Sterne werden gesetzt?

Im Baum T verzweigen wir zuerst nach einer zufälligen Restriktion ρ und verfeinern ρ mit Hilfe von T_ρ.

Wir setzen $t = 2\log_2 g$ und wählen $l = \Theta(n/t)$ Sterne.

- Wir fordern, dass nicht zu viele gesternte Variablen in T gesetzt sind, um weitere Schichten von S mit neuen Restriktionen kippen zu können.
- Genau diese Forderung erreichen wir
  mit dem Switching Lemma für eine ganze Schicht,
  wenn wir $s = \alpha n/t$ für eine Konstante α setzen:

  Die Wahrscheinlichkeit, dass $\geqslant s$ gesternte Variablen in T gesetzt werden, ist höchstens $2^{-O(n/t)}$.

T besitzt also höchstens

$$1 \cdot 2^{n-O(n/t)} + 2^{-O(n/t)} 2^n = 2^{n-O(n/t)}$$

Blätter.

Zusammenfassung

Wende eine neue Restriktion auf die $\Theta(\frac{n}{t})$ gesternten Variablen an.

- Wir arbeiten wieder mit Bottom-Fanin $t = 2\log_2 g$ und wählen diesmal $\frac{n}{t^2}$ Sterne: Es gibt höchstens $2^{n-O(n/t^2)}$ Blätter.
- Nach $g-2$ Wiederholungen, haben wir höchstens $2^{n-O(n/t^{g-2})}$ Blätter im entsprechenden Entscheidungsbaum.
  - **Aber** der Weg zu einem Blatt wird durch eine KNF mit Bottom-Fanin t und eine Restriktion beschrieben.
  - Ebenso ist der Schaltkreis auf dem Weg zum Blatt nicht konstant, sondern ist (o.B.d.A.) äquivalent zu einer KNF mit Bottom-Fanin t.
  - Wir machen beide KNFs mit einer konventionellen Restriktion platt:

Sei S ein Schaltkreis der Tiefe d, Größe g und mit Bottom-Fanin $t = 2\log_2 g$. Dann gibt es eine Zerlegung des n-dimensionalen Würfels in höchstens

$$2^{n-n/O(\log_2 g)^{d-2}}$$

Teilwürfel, so dass S auf jedem Teilwürfel konstant ist. (S hat n Eingabebits.)

Korrelation von Schaltkreisen und Paritätsfunktion

Korrelation

Die Korrelation einer Booleschen Funktion $f : \{0,1\}^n \to \{0,1\}$ mit einer Booleschen Funktion $g : \{0,1\}^n \to \{0,1\}$ ist genau dann c, wenn

$$\mathrm{pr}[\, f(x) = g(x) \,] = \frac{1}{2} + c.$$

- Wenn der Schaltkreis S auf einem Teilwürfel T der Dimension größer als Eins konstant ist, dann ist die Korrelation von S mit $\mathtt{xor}_n$ auf T Null.
- Bestenfalls besitzen alle Teilwürfel, bis auf einen, die Dimension 1 und $2^{n-n/O(\log_2 g)^{d-2}}/2^n$ ist die (durchschnittliche) Korrelation.

Sei S ein Schaltkreis der Tiefe d und Größe g mit Bottom-Fanin $2\log_2 g$. Die Korrelation von S mit $\mathtt{xor}_n$ ist $\leqslant 2^{-n/O(\log_2 g)^{d-2}}$, d.h. es gilt

$$\mathrm{pr}[\, S(x) = \mathtt{xor}_n(x) \,] \leqslant \frac{1}{2} + 2^{-n/O(\log_2 g)^{d-2}}.$$

$\{\neg, \wedge, \vee, \mod p\}$-Schaltkreise

p sei eine Primzahl.

Bleibt die Paritätsfunktion schwierig?

- Die Methode der Restriktionen versagt, denn ein mod p-Gatter lebt, wenn nicht alle Variablen gesetzt werden.
- Razborov-Smolensky führen eine neue, algebraische Methode ein: Der $\{\neg, \wedge, \vee, \mod p\}$-Schaltkreis S wird durch ein Polynom **approximiert**.
  - Wenn S nicht zu groß ist und eine kleine Tiefe hat, dann hat das Polynom einen relativ kleinen Grad.
  - Man zeigt, dass die Paritätsfunktion nur von Polynomen mit großem Grad erfolgreich approximiert werden kann.

Das Ergebnis

Die Paritätsfunktion hat nur $\{\neg, \wedge, \vee, \mod p\}$-Schaltkreise der Tiefe d, wenn die Größe mindestens

$$2^{n^{1/2d}}$$

beträgt.

$\{\neg, \wedge, \vee, \mathrm{mod}\ m\}$-Schaltkreise

m sei eine beliebige möglicherweise **zusammengesetzte** Zahl.

$\mathrm{AC}^0[m]$ ist die Klasse aller Sprachen, die durch Schaltkreisfamilien beschränkter Tiefe und polynomieller Größe berechenbar sind, wenn neben den $\neg, \wedge$ und $\vee$-Gattern noch mod_m Gatter benutzt werden können.

Es ist noch nicht einmal bekannt, ob $\mathrm{AC}^0[m] = \mathrm{NP}$ gilt!