

Tweede Hertoets Concurrency

24 maart 2016, 11.00–13.00, Unnik-220.

Motiveer je antwoorden *kort!* Zet je mobiel uit. Stel geen vragen over deze toets; als je een vraag niet duidelijk vindt, schrijf dan op hoe je de vraag interpreteert en beantwoord de vraag zoals je hem begrijpt.

Cijfer: Maak vraag 1 en 2 op pagina 1, vraag 3, 4 en 5 op pagina 2, en vraag 6 en 7 op pagina 3. Vragen 3 t/m 6 zijn elk 2pt, vragen 1, 2 en 7 zijn 3pt. Cijfer is totaal plus 1, gedeeld door 1,7.

1. **Monte Carlo Half:** Hugo wil de waarde van een half numeriek benaderen door N keer twee random doubles x en y te trekken, in variabele S te tellen hoe vaak $x < y$ geldt, en dan $h = S/N$ af te drukken.

(a) Beredeneer met behulp van Variantie, hoe groot hij N moet maken om een half op 5 decimalen nauwkeurig uit te rekenen.

(b) Om minder random getallen te hoeven trekken, wil Hugo de y van elk experiment hergebruiken als de x van het volgende experiment. Bespreek of dit een goed idee is.

Oplossing: (a) Van een enkele trekking is de slaagkans p een half, de faalkans q dus ook, dus de variantie is $V[X] = p \cdot q = 1/4$ en $\sigma[X] = 1/2$. De standaarddeviatie van output h is dus $\sigma(h) = \frac{1/2}{\sqrt{N}}$. Deze σ is kleiner dan 0,00001 zodra $N > (\frac{1/2}{0,00001})^2 = 2.500.000.000$. Met nog viermaal zoveel samples is de standaarddeviatie zelfs maar 0,000005 zodat de kans 95% is dat het resultaat binnen 0,00001 van het juiste antwoord ligt.

(b) De opeenvolgende experimenten zijn niet langer onafhankelijk. Voor een getal y dat bv. *groter* was dan x , is de kans niet een half maar eenderde dat het vervolgens kleiner is dan het volgende getrokken getal. Omdat de variantie-analyse van Monte Carlo simulaties op onafhankelijkheid is gebaseerd, begeeft Hugo zich op een glibberig pad, waarvoor hij beter eerst nog wat aanvullende colleges kan volgen.

Beoordeling/Toelichting: Totaal 3pt, 2pt voor(a) en 1pt voor (b).

2. **Bayes: Rood en Geel:** Er zijn twee zakken die op elkaar lijken: zak Z1 met 10 rode en 5 gele ballen, zak Z2 met 10 gele en 2 rode ballen. Je pakt willekeurig een zak en daaruit willekeurig een bal; deze is rood.

(a) Hoe luidt de Regel van Bayes?

(b) Wat is de kans dat de gepakte zak, zak Z1 is?

Oplossing: (a) $P(E|F) = P(F|E) \cdot \frac{P(E)}{P(F)}$.

(b) Die is 4/5.

Beoordeling/Toelichting: Max 3, 1 voor de Regel en 2pt voor de toepassing.

K = Dat Klopt wel, maar 't is niet de Regel van Bayes.

3. **SIMD/Reductie:** Geef schematisch weer hoe een input set van 4 elementen a_0 t/m a_3 efficiënt met 2-wide SIMD gereduceerd kan worden tot een enkele waarde door herhaalde toepassing van een binaire operatie. Leg uit wat in dit geval de eigenschappen moeten zijn van de binaire operatie, en waarom.

Oplossing: Schema: $r = a_0 x a_1 x a_2 x a_3 = ((a_0 x a_2) x (a_1 x a_3))$. a_2 en a_1 zijn van plaats verwisseld; de operatie moet dus zowel commutatief als associatief zijn.

Beoordeling/Toelichting: Voor correct schema 1 punt. Voor voorwaarden en uitleg 1 punt.

4. **GPU hardware:** Waarom is op een GPU de cache minder belangrijk dan op een CPU?

Oplossing: De GPU gebruikt hyperthreading op veel grotere schaal. Waar de CPU caches gebruikt om latencies te verlagen worden deze latencies op de GPU verborgen door ander werk te gaan doen. Dit vereist een hoge mate van parallelisme.

Beoordeling/Toelichting: Twee punten voor een correct antwoord.

5. **Associativiteit:** Laat met een voorbeeld zien dat floating point addition niet associatief is.

Oplossing: Hier zijn zeer veel mogelijkheden; problemen treden vooral op wanneer de drie getallen van verschillende grootte orde zijn.

Beoordeling/Toelichting: Twee punten voor een correct voorbeeld.

6. **Begrippen:** Wat is het verschil tussen:

- (a) SIMD en SIMT?
- (b) Stencil en map?

Oplossing: (a) Het verschil is wat subtiel. Met SIMD bieden we het algoritme expliciet vectoren aan. Met SIMT bieden we groepen threads aan; de data die verwerkt wordt in deze threads is georganiseerd in vectoren. Er is dus sprake van implicit SIMD.

(b) Een stencil gebruikt meerdere elementen als input, waarbij overlap optreedt tussen kernel inputs. Bij een map treedt deze overlap niet op.

Beoordeling/Toelichting: Voor elk correct antwoord een punt.

7. **Vectorizatie:** Twee vormen van parallelisme zijn instruction level parallelism en thread level parallelism.

- (a) Wat is het verschil tussen beide vormen?
- (b) Wat is het verschil tussen een thread en een stream (of lane)?
- (c) Op welke manier past een GPU thread level en instruction level parallelism toe?

Oplossing: (a) Threads worden onafhankelijk van elkaar uitgevoerd, al dan niet op verschillende cores. Instruction level parallelism kan in een enkele thread worden uitgevoerd; de threads lopen dan in lock-step en moeten hetzelfde algoritme uitvoeren, eventueel met masking (speculative execution) om verschillen in flow af te handelen. Heel kort: threads hebben elk een eigen program counter, streams niet.

(b) Threads hebben elk een eigen program counter, streams niet. Streams worden in lock-step uitgevoerd.

(c) Dit is een inzichtvraag. Threads binnen een warp worden in lockstep uitgevoerd; dit is dus impliciet instruction level parallelism. Warps als geheel gedragen zich als threads, die met hyperthreading op een of meerdere GPU cores uitgevoerd worden.

Beoordeling/Toelichting: Voor elk correct antwoord een punt.