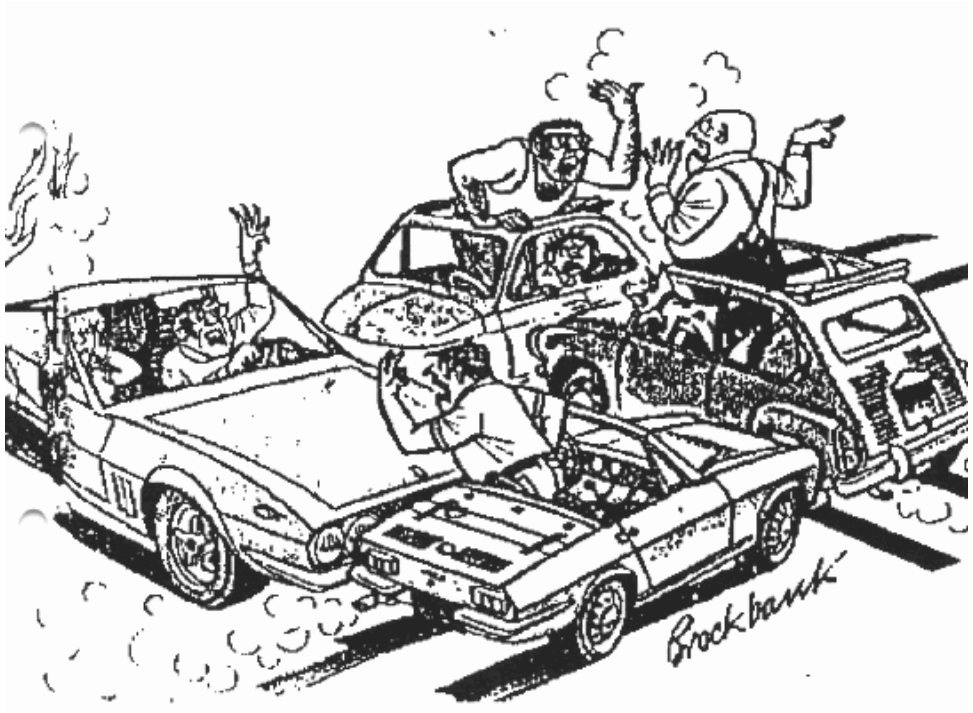
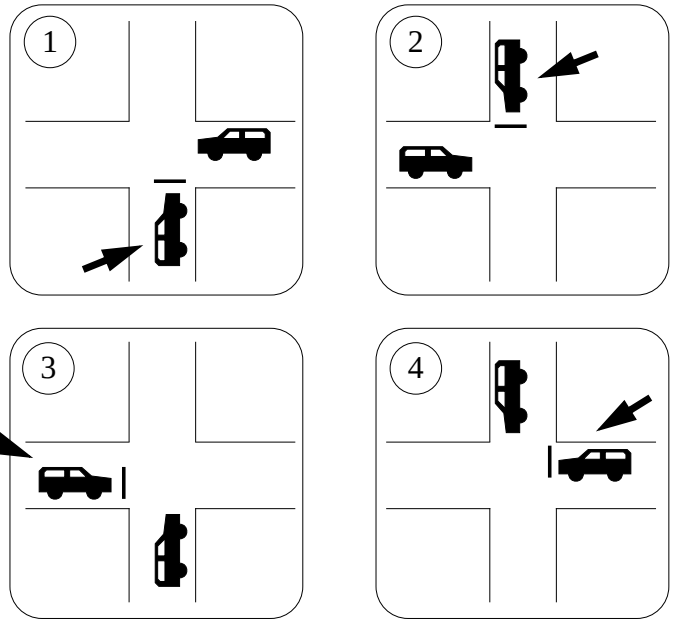
Ein zweites Beispiel: Das Deadlock-Problem



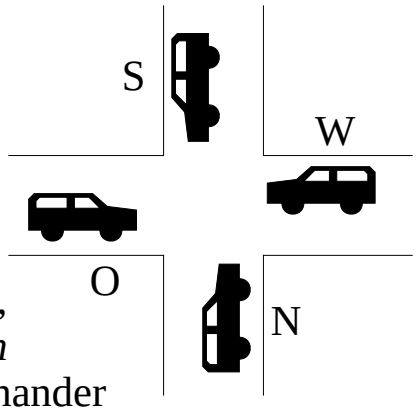
„Neapolitanischer Hakenkreuzstau“
(Also sprach Bellavista, Luciano De Crescenzo)

Phantom-Deadlocks



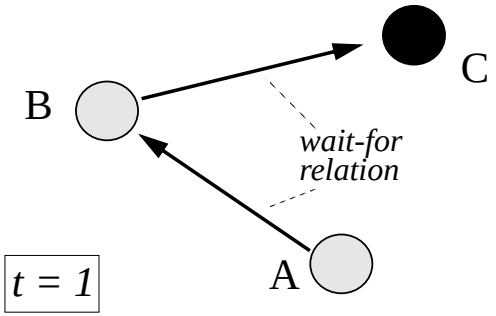
Vier Einzelbeobachtungen der Autos N, S, O, W

1) N wartet auf W
2) S wartet auf O
3) O wartet auf N
4) W wartet auf S



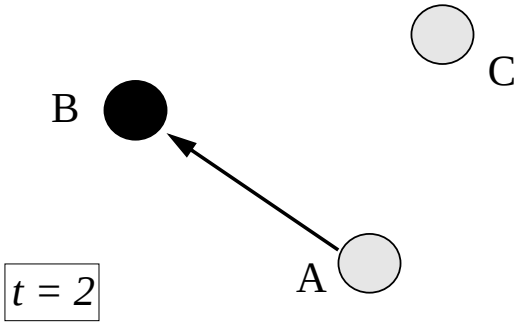
zu notwendigerweise verschiedenen Zeitpunkten liefert den *falschen* Eindruck, als würden zu einem *einzigen* Zeitpunkt alle zyklisch aufeinander warten (--> Verklemmung)

Phantom-Deadlocks

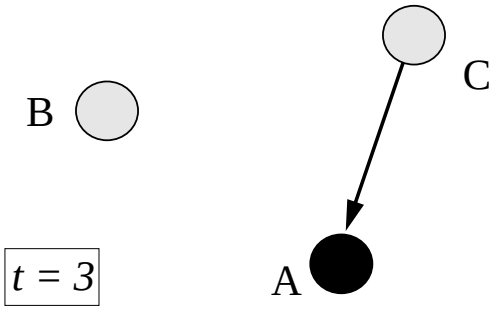


(C benutzt ein exklusives Betriebsmittel)

beobachte B:
==> B wartet auf C



beobachte A:
==> A wartet auf B

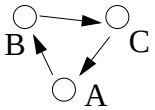


beobachte C:
==> C wartet auf A

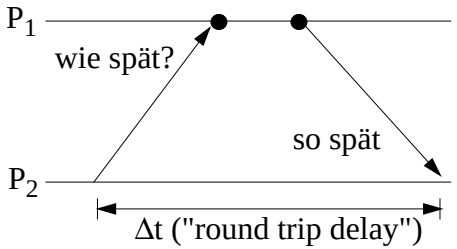
falscher Schluss!

Deadlock!

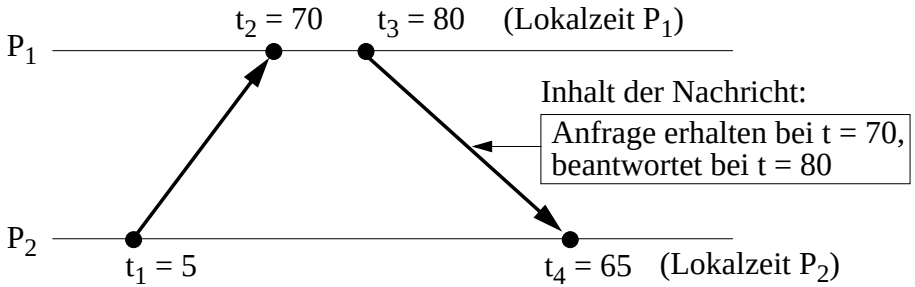
Keine exakte globale Zeit!



Ein drittes Problem: Uhrensynchronisation



- Lastabhängige Laufzeiten von Nachrichten
- Unsymmetrische Laufzeiten
- Wie erfährt man die Laufzeit?

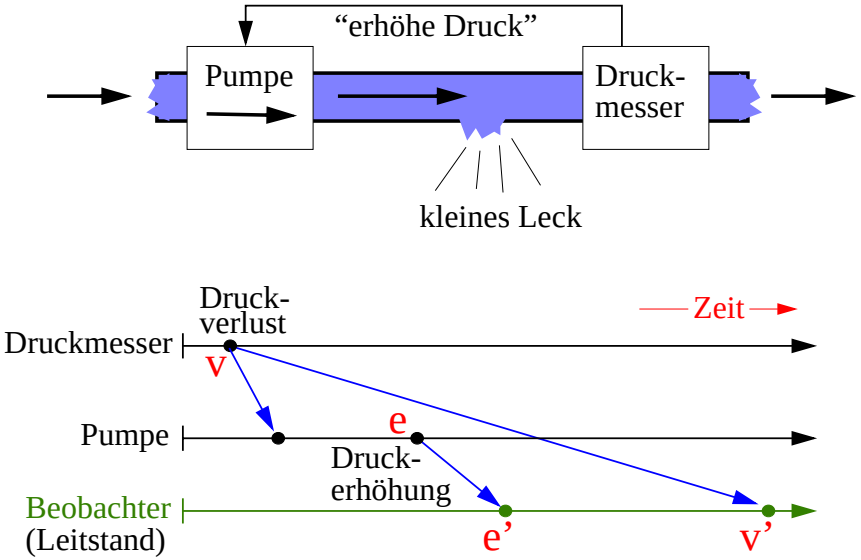


- Uhren gehen nicht unbedingt gleich schnell!
  (wenigstens “Beschleunigung ≈ 0”, d.h. konstanter Drift gerechtfertigt?)

- Wie kann man den Offset der Uhren ermitteln oder zumindest approximieren?

Ein viertes Problem: Kausal (in)konsistente Beobachtungen

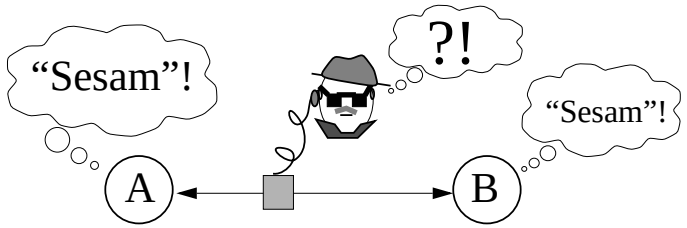
- Gewünscht: Eine Ursache stets vor ihrer (u.U. indirekter) Wirkung beobachten



Falsche Schlussfolgerung des Beobachters:

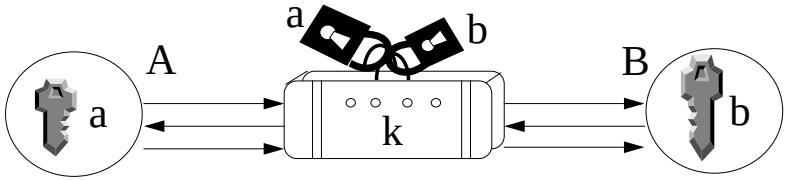
Eine unbegründete Pumpenaktivität erhöhte den Druck bis zum Bersten der Pipeline; daraufhin trat das Öl aus dem Leck aus, was durch den Druckverlust angezeigt wird!

Und noch ein Problem: Verteilte Geheimnisvereinbarung



- Problem: A und B wollen sich über einen unsicheren Kanal auf ein gemeinsames geheimes Passwort einigen.

- Idee: Vorhängeschlösser um eine sichere Truhe:



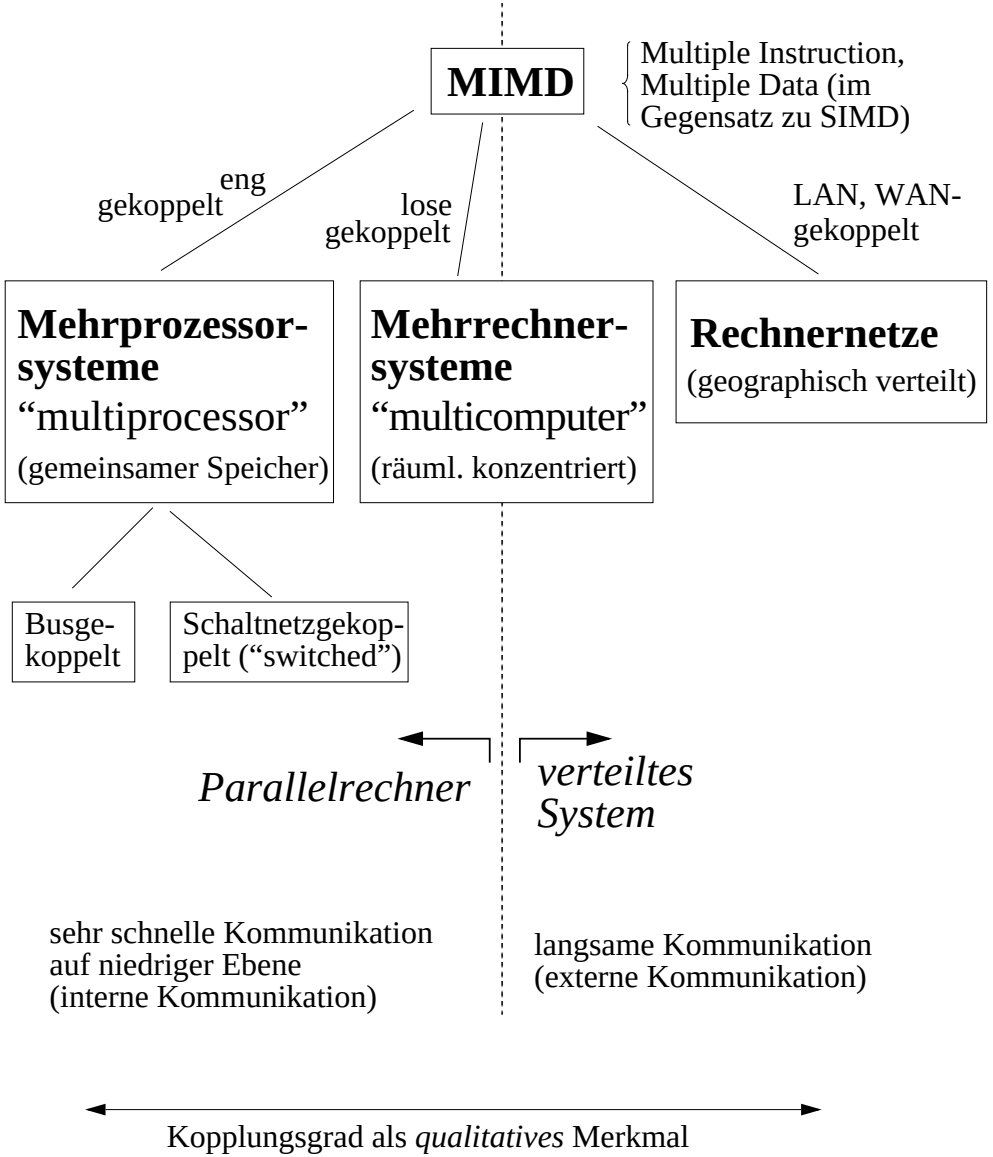
1. A denkt sich Passwort k aus und tut es in die Truhe.
2. A verschliesst die Truhe mit einem Schloss a.
3. A sendet die so verschlossene Truhe an B.
4. B umschliesst das ganze mit seinem Schloss b.
5. B sendet alles doppelt verschlossen an A zurück.
6. A entfernt Schloss a.
7. A sendet die mit b verschlossene Truhe wieder an B.
8. B entfernt sein Schloss b.

- Problem: Lässt sich das so softwaretechnisch realisieren?

  Wie wäre es damit?: k sei eine Zahl. “Verschliessen” und “aufschliessen” eines Schlosses entspricht dem Hinzuaddieren oder Subtrahieren einer beliebig ausgedachten (geheimgehaltenen) Zahl a bzw. b.

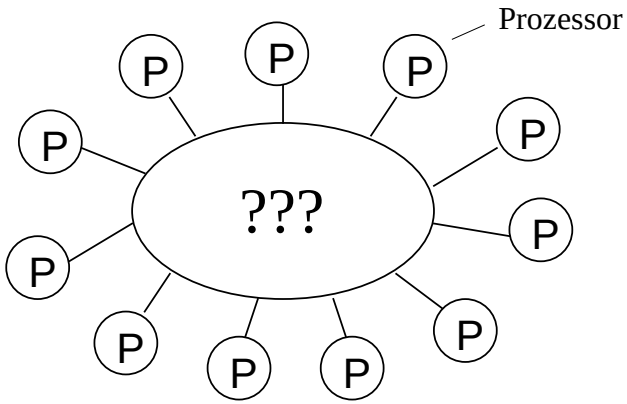
Multiprozessoren und Multicomputer

Abgrenzung Parallelrechner



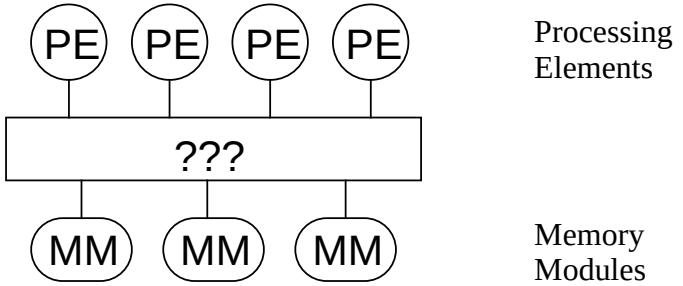
Prozessorverbund

- Autonome Prozessoren + „Kommunikationsnetz“
- Je nach Kopplungsgrad und Grad der Autonomie ergibt sich daraus ein
  - Mehrprozessorsystem
  - Mehrrechnersystem
  - Rechnernetz



Speicherkopplung

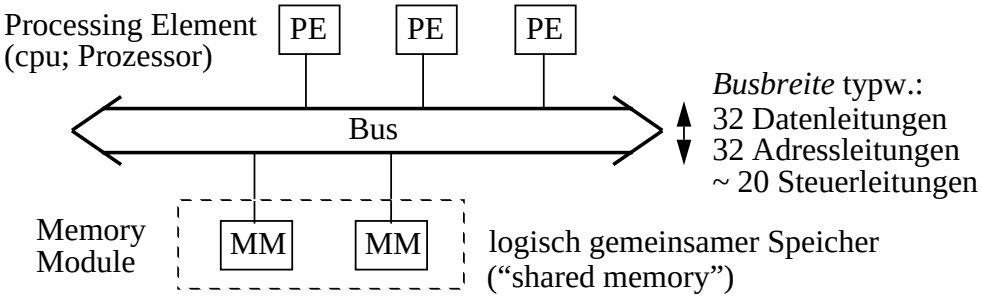
- Shared Memory
  - Kommunikation über gemeinsamen Speicher



- n Processing Elements teilen sich k Memory Modules
- Kopplung zwischen PE und MM, z.B.
  - Bus
  - Schaltnetz
  - Permutationsnetz
- UMA-Architektur (Uniform Memory Access) oder NUMA (Non-Uniform Memory Access)

wenn es “nahe” und “ferne” Speicher gibt: z.B. schneller Zugriff auf den “eigenen” Speicher, langsamer auf fremden

Busgekoppelte Multiprozessoren

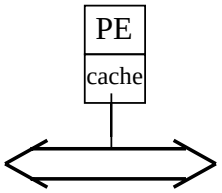


Problem:

Bus i.a. bereits bei wenigen (3 - 5) PEs überlastet

Lösung:

Lokale Caches
zwischen PE und Bus:



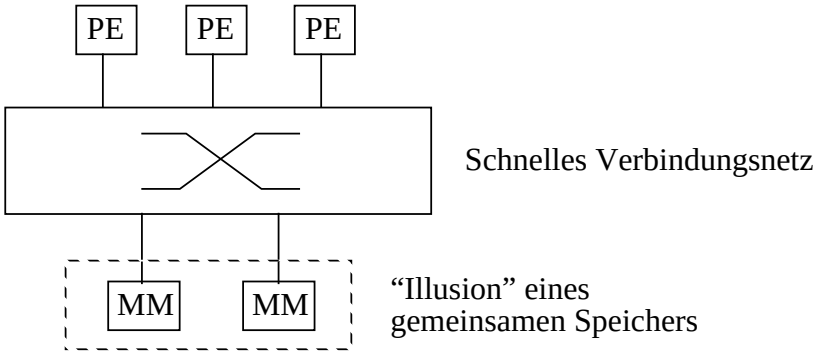
Cache gross genug wählen, um Hitraten > 90% zu erzielen (abhängig von der Hauptspeichergrösse)!

Probleme:

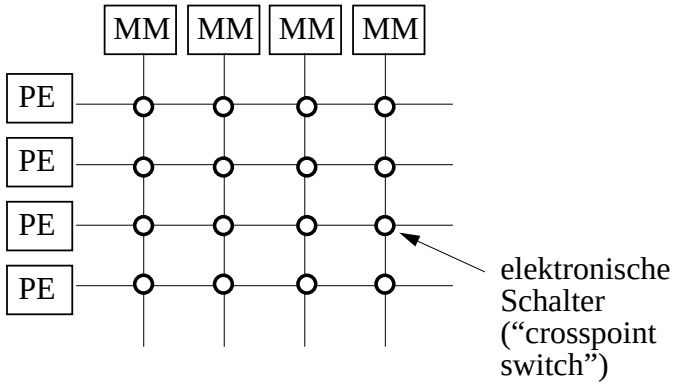
1) Kohärenzproblem der caches
2) Damit Problem nur verschoben
(ca. 10 Mal mehr Prozessoren möglich)

Generell: Busgekoppelte Systeme schlecht skalierbar!
(Übertragungsbandbreite bleibt “konstant” bei Erweiterung um Knoten)

Schaltnetzgekoppelte Multiprozessoren



Z.B. Crossbar-switch (Kreuzschienenverteiler):

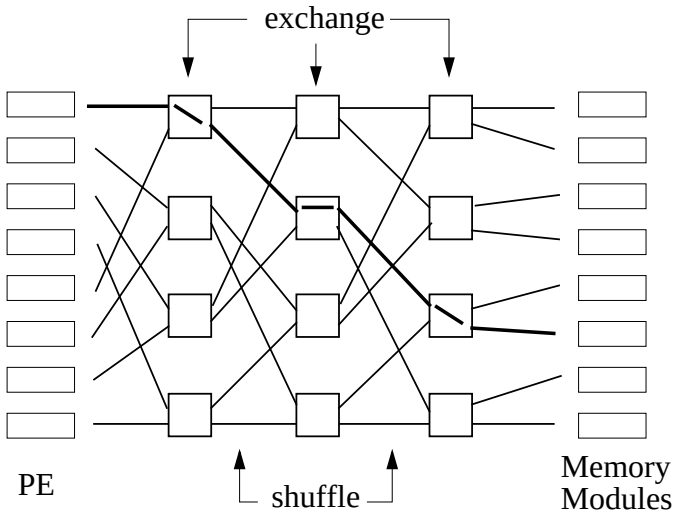


- Mehrere PEs können gleichzeitig auf verschiedene Speichermodule zugreifen
- Schlecht skalierbar (quadratisch viele Schalter)
(Vermeidung von hot spots durch interleaving, Randomisierung...)

Permutationsnetze

Mehrere Stufen von Schaltelementen ermöglichen die Verbindung jedes Einganges zu jedem Ausgang.

Schaltelement (“interchange box”) kann zwei Zustände annehmen (durch ein Bit ansteuerbar):



Beispiel:

*Shuffle-Exchange-*Netz

(Omega-Netz)

Hier: log n (identische!) Stufen mit je n/2 Schaltern.

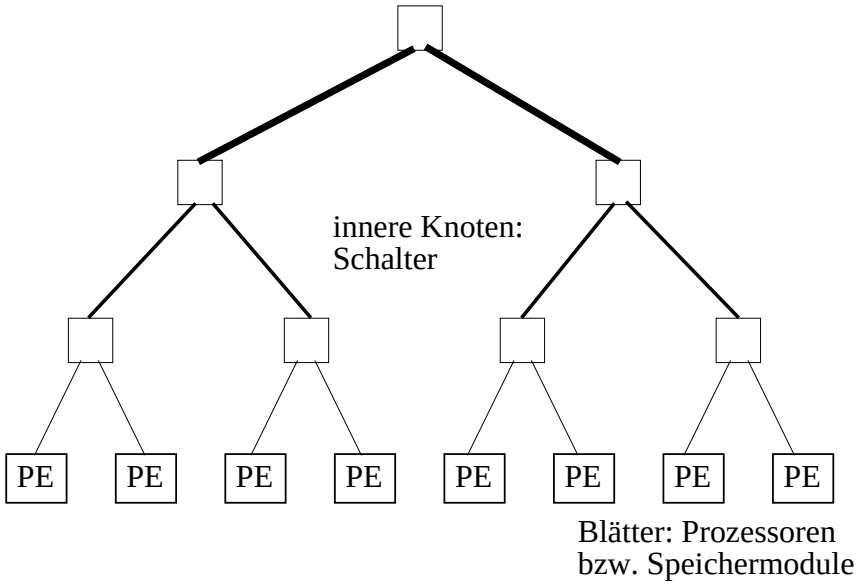
Es gibt weitere ähnliche dynamisch schaltbare Netze.

Designkriterien:

z.B. Butterfly-Netze

- wenig Stufen (“delay”)
- Parallele Zugriffe; Vermeidung von Blockaden
- ggf. Zugriffe bündeln (“combining”)

Fat-Tree-Netze



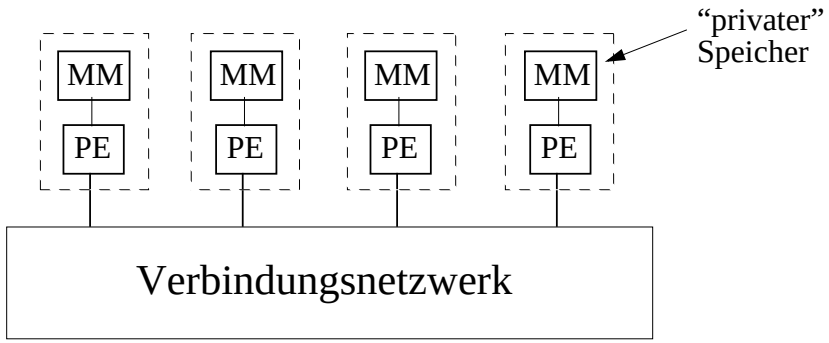
Verbindungsleitungen höherer Bandbreite bzw. mehrere parallele Leitungen auf Niveaus, die näher an der Wurzel liegen.

Multiprozessoren — Fazit

- Gemeinsamer Speicher, über den die Prozessoren Information austauschen (d.h. kommunizieren) können
  - Prozessoren müssen mit dem Speicher (bzw. den einzelnen Speichermodulen) gekoppelt werden
- Speicherkopplung begrenzt Skalierbarkeit und räumliche Ausdehnung
  - Untergliederung des Speichers in mehrere Module (Parallelität)
  - leistungsfähiges Kommunikationsnetz
- Lokale PE-Caches sinnvoll
  - Problem der Cache-Kohärenz
- Bewertungskriterien für Verbindungsnetze
  - Realisierungsaufwand (Fläche, Kosten)
  - Skalierbarkeit (mit wachsender Anzahl PEs und MMs)
  - innere Blockadefreiheit (parallele Kommunikationsvorgänge)
  - Anzahl der Stufen (Verzögerung)
  - Eingangsgrad, Ausgangsgrad der Bauelemente

Mehrrechnersysteme ("Multicomputer")

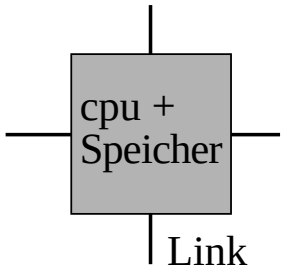
Vernetzung vollständiger Einzelrechner:



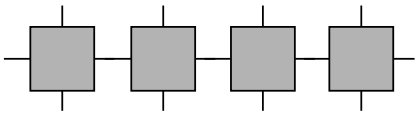
Zugriff auf andere Rechner (bzw. deren private Speicher) nur indirekt über *Nachrichten*.

- kein globaler Speicher
- NORMA-Architektur (NO Remote Memory Access)

Beispiel: Transputer als Baustein für Multicomputer

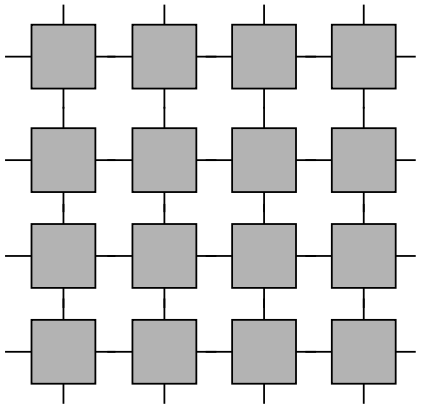


- Typische Topologien:



Pipeline

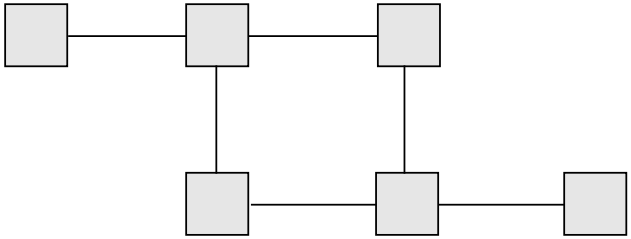
Array (oder Torus)



- bidirektional über mehrere Meter
- kleine set-up time
- synchrone Komm.
- Mehrprozesskonzept mit Scheduler "on chip"
- Eigene Program-miersprache "Occam"

- Hersteller: INMOS (GB)
- Erste Modelle: 1983; T414 (1986), T800 (1987)
  - in den 90er-Jahren keine Nachfolgetypen mehr entwickelt

Verbindungstopologien für Mehrrechnersysteme



Zusammenhängender Graph mit

Knoten = Rechner

Kante = dedizierte Kommunikationsleitung

Ausdehnung: i.a. nur wenige Meter

Bewertungskriterien:

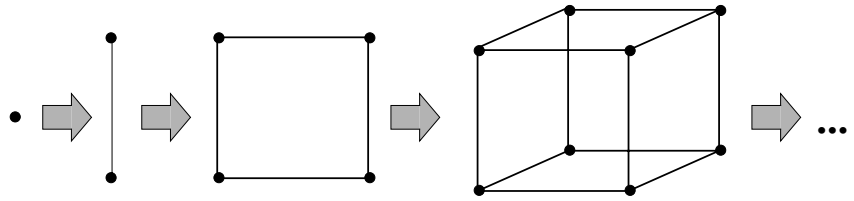
- Gesamtzahl der Verbindungen (bei n Knoten)
- maximale Entfernung zweier Knoten
- durchschnittliche Entfernung
- Anzahl der Nachbarn eines Knotens ("fan out")
- Symmetrie, Homogenität, Skalierbarkeit...
- Routingkomplexität
- Zahl der alternativ bzw. parallel verfügbaren Wege

Technologische Faktoren:

- Geschwindigkeit, Durchsatz, Verzögerung, eigene Kommunikationsprozessoren...

Hypercube

- Hypercube = "Würfel der Dimension d"



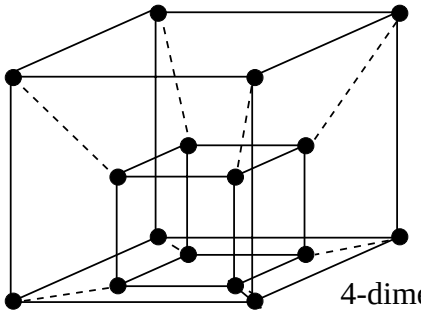
Draufsicht von der Seite liefert jeweils niedrigere Dimension

Entsprechend: Herausdrehen des Objektes aus der Blickebene zeigt, dass es sich "eigentlich" um ein Objekt der Dimension n+1 handelt!

- Rekursives Konstruktionsprinzip

  - Hypercube der Dimension 0: Einzelrechner
  - Hypercube der Dimension d+1:

    „Nimm zwei Würfel der Dimension d und verbinde korrespondierende Ecken"



4-dimensionaler Würfel

Man vgl. auch das Buch von T. F. Banchoff: Beyond the Third Dimension (Scientific American Library, 1990)

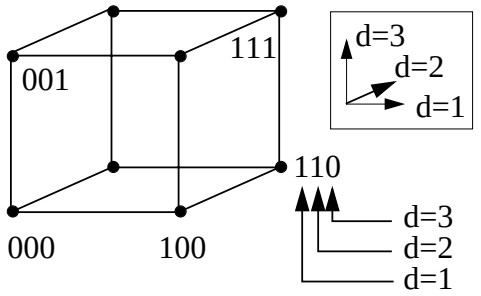
Hypercube der Dimension d

- $n = 2^d$ Knoten

- Anzahl der Nachbarn eines Knotens = d
  (Anzahl der "ports" in der Hardware)

- Gesamtzahl der Kanten (= Verbindungen): $d\ 2^d/2 = d\ 2^{d-1}$
  (Ordnung O(n log n))

- Einfaches Routing:
  - Knoten systematisch (entspr. rekursivem Aufbau) numerieren
  - Zieladresse bitweise xor mit Absenderadresse
  - Wo sich eine "1" findet, in diese Dimension muss gewechselt werden



- Maximale Weglänge: d

- Durchschnittliche Weglänge = d/2
  (Induktionsbeweis als Übung!)

-Vorteile Hypercube:

- kurze Weglängen (max. log n)
- einfaches Routing
- viele Wegalternativen (Fehlertoleranz, Parallelität!)

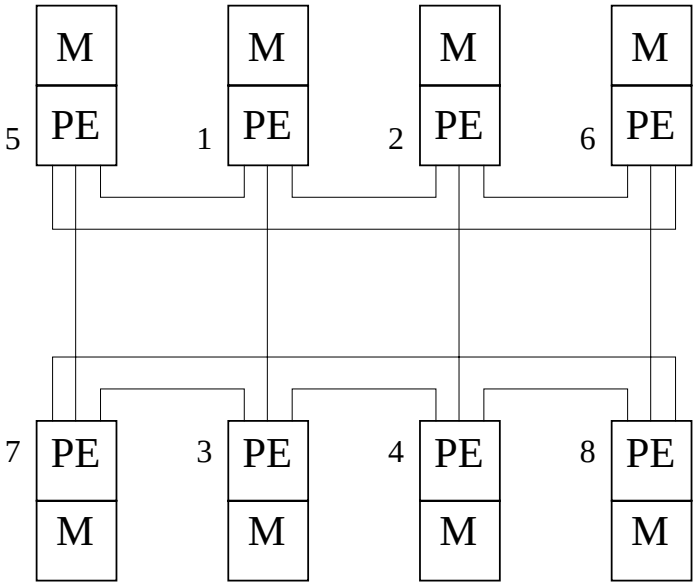
Denkübung: mittlere Weglänge?

-Nachteile:

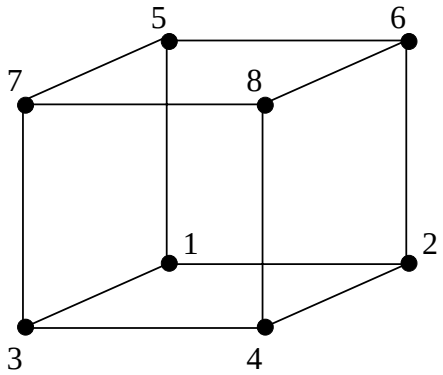
- Anzahl der Nachbarknoten eines Knotens wächst mit der Dimension d
- insgesamt relativ viele Verbindungen: O(n log n)
  (eigentlich genügen n-1!)

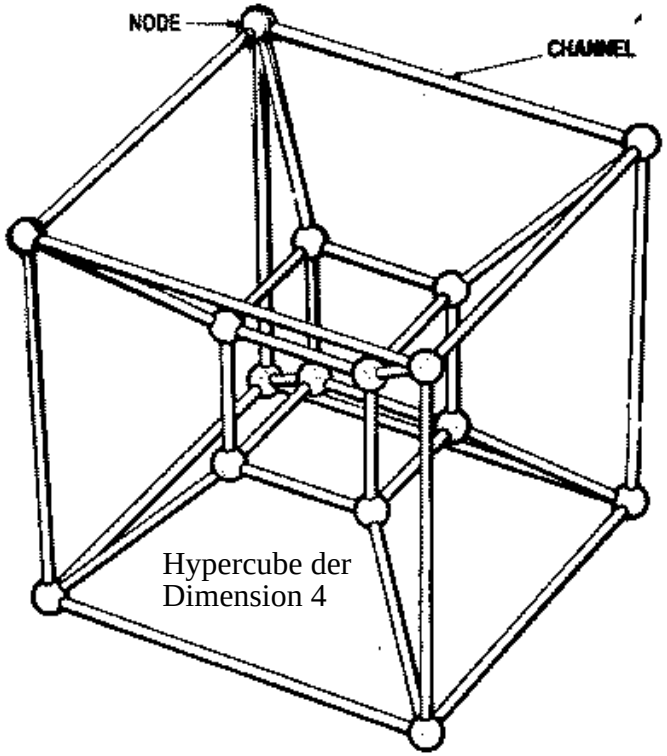
wieviele verschiedene Wege der Länge k gibt es insgesamt?

Layout eines Hypercube



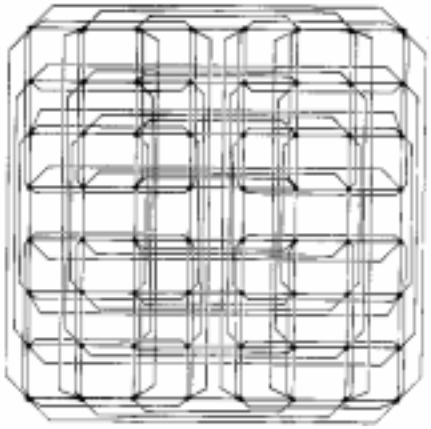
Obiger Topologie sieht man zunächst nicht an, dass es sich dabei um einen 3-dimensionalen Würfel handelt!





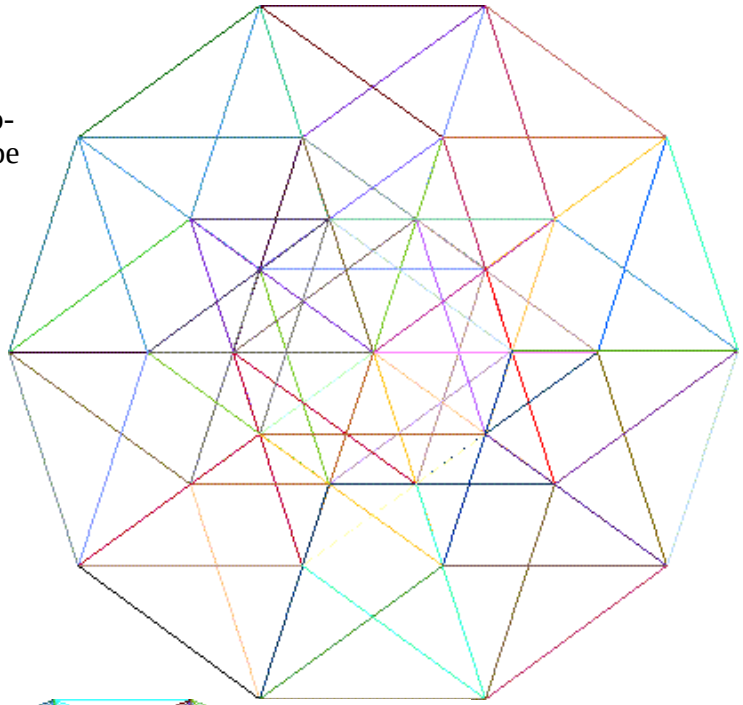
Hypercube der Dimension 4

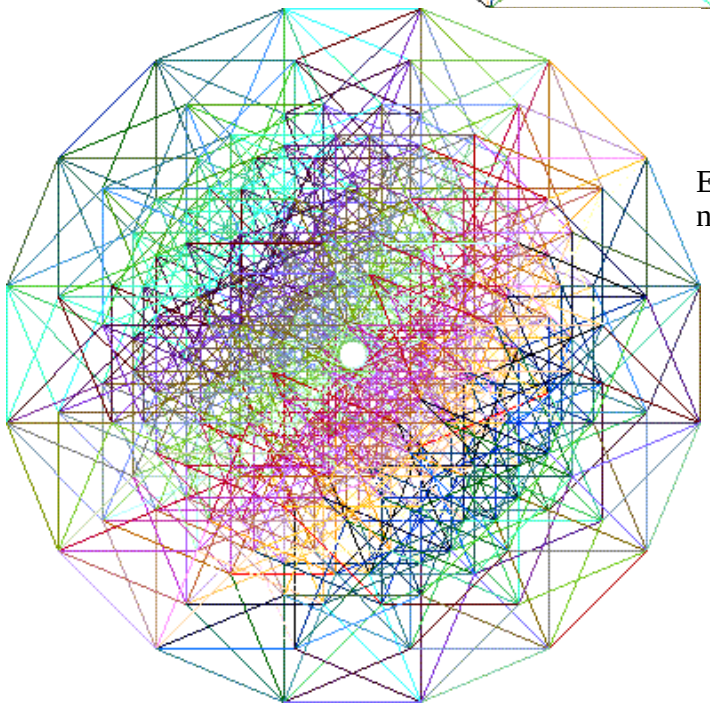
Hypercube der Dimension 6 in der Ebene



Ein 5-dimensionaler Hypercube

www.cs.reading.ac.uk/archive/hypercubes/





Ein 8-dimensionaler Hypercube

www.comlab.ox.ac.uk/oucl/users/jonathan.bowen/publications/ndcubes.html

Broadcast in Hypercubes

- Broadcast: Eine Nachricht an alle anderen Knoten senden.
- Initiator habe die Nummer 00...00 (binär).
- Initiator sendet an alle seine Nachbarn:

  0...01, 0...010, 0...100, ... , 10...0

  in “kanonischer” Numerierung

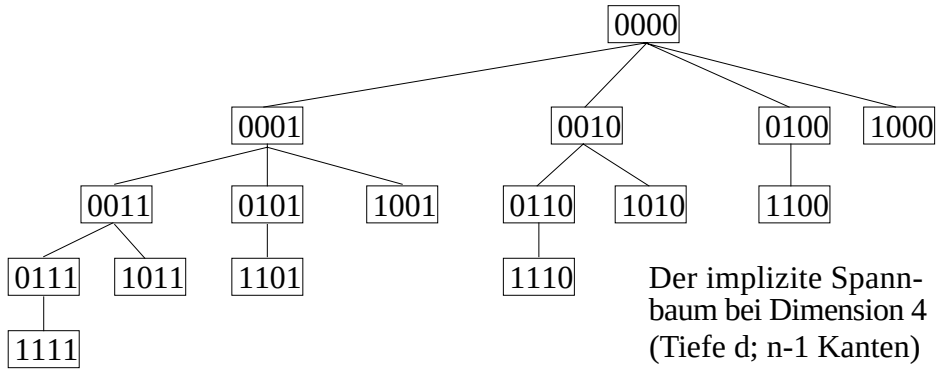
  am besten gleichzeitig, wenn dies technisch geht!

- Ein Knoten mit der Nummer 0...01x...y...z leitet die Information an alle seine “höheren” Nachbarn weiter:

  linkeste 1 | beliebiges Restmuster

  0...00**1**1x...y...z
  0...0**1**01x...y...z
  0...**1**001x...y...z
  ...
  10...001x...y...z

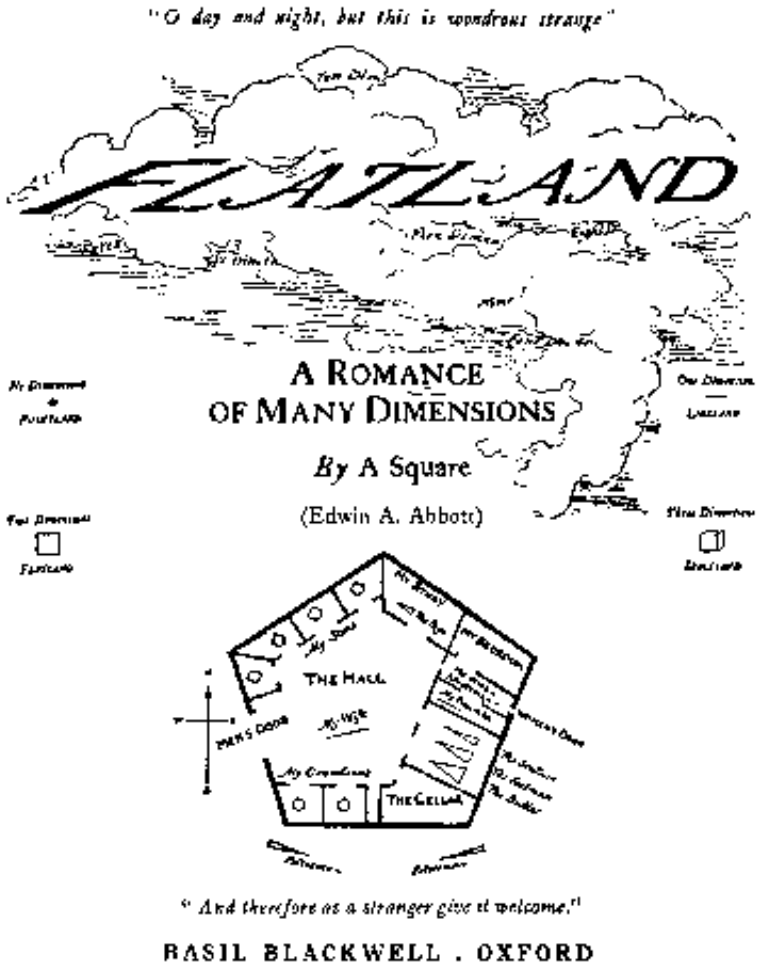
  Von welchem (eindeutigen) Knoten A wird Knoten B informiert?
  Setze *vorderste 1* von B auf 0
  --> = Nummer von A



Der implizite Spannbaum bei Dimension 4 (Tiefe d; n-1 Kanten)

- Der Algorithmus wird z.B. in Mehrprozessorsystemen (z.B. NCube) verwendet
- Wie effizient ist der Algorithmus? (Geht es besser?)
- Denkübung: Formuliere Algorithmus für einen beliebigen Initiator (schliesslich sind Hypercubes symmetrisch...)
- Andere (bessere?) Algorithmen in der Vorlesung “Verteilte Algorithmen”

Flatland (1884)



- Diskussion über Hypercubes und höhere geometrische Dimensionen zwischen “A. Square” und “Sphere”

- Gleichzeitig soziale Satire

§ 3 .---*Concerning the Inhabitants of Flatland*

...

Our Women are Straight Lines.

Our Soldiers and Lowest Class of Workmen are Triangles with two equal sides, each about eleven inches long...

Our Middle Class consists of Equilateral or Equal-Sided Triangles.

Our Professional Men and Gentlemen are Squares (to which class I myself belong) and Five-Sided Figures or Pentagons.

Next above these come the Nobility, of whom there are several degrees, beginning at Six-Sided Figures, or Hexagons, and from thence rising in the number of their sides till they receive the honourable title of Polygonal, or many-Sided. Finally when the number of the sides becomes so numerous, and the sides themselves so small, that the figure cannot be distinguished from a circle, he is included in the Circular or Priestly order; and this is the highest class of all.

...

§ 4 .---*Concerning the Women*

If our highly pointed Triangles of the Soldier class are formidable, it may be readily inferred that far more formidable are our Women.

...

But here, perhaps, some of my younger Readers may ask HOW a woman in Flatland can make herself invisible. This ought, I think, to be apparent without any explanation. However, a few words will make it clear to the most unreflecting.

Place a needle on the table. Then, with your eye on the level of the table, look at it side-ways, and you see the whole length of it; but look at it end-ways, and you see nothing but a point, it has become practically invisible. Just so is it with one of our Women.

...

The dangers to which we are exposed from our Women must now be manifest to the meanest capacity of Spaceland. If even the angle of a respectable Triangle in the middle class is not without its dangers... --what can it be to run against a woman, except absolute and immediate destruction? And when a Woman is invisible, or visible only as a dim sub-lustrous point, how difficult must it be, even for the most cautious, always to avoid collision!

...

In the Southern and less temperate climates, where the force of gravitation is greater, and human beings more liable to casual and involuntary motions, the Laws concerning Women are naturally much more stringent. But a general view of the Code may be obtained from the following summary:--

1. Every house shall have one entrance on the Eastern side, for the use of Females only; by which all females shall enter "in a becoming and respectful manner" and not by the Men's or Western door.

2. No Female shall walk in any public place without continually keeping up her Peace-cry, under penalty of death.

3. ...

In some of the States there is an additional Law forbidding Females, under penalty of death, from walking or standing in any public place without moving their backs constantly from right to left so as to indicate their presence to those behind them...

§ 16 .---*How the Stranger vainly endeavoured to reveal to me in words the mysteries of Spaceland*

...

Sphere. ... We began with a single Point, which of course -- being itself a Point -- has only ONE terminal Point. One Point produces a Line with TWO terminal Points.

One Line produces a Square with FOUR terminal Points.

Now you can give yourself the answer to your own question: 1, 2, 4, are evidently in Geometrical Progression. What is the next number?

I. Eight.

Sphere. Exactly. The one Square produces a SOMETHING-WHICH-YOU-DO-NOT-AS-YET-KNOW-A-NAME-FOR-BUT-WHICH-WE-CALL-A-CUBE with EIGHT terminal Points. Now are you convinced?

...

Sphere. How can you ask? And you a mathematician! The side of anything is always, if I may so say, one Dimension behind the thing. Consequently, as there is no Dimension behind a Point, a Point has 0 sides; a Line, if I may so say, has 2 sides (for the points of a Line may be called by courtesy, its sides); a Square has 4 sides; 0, 2, 4; what Progression do you call that?

I. Arithmetical.

Sphere. And what is the next number?

I. Six.

Sphere. Exactly. Then you see you have answered your own question. The Cube which you will generate will be bounded by six sides,that is to say, six of your insides. You see it all now, eh?

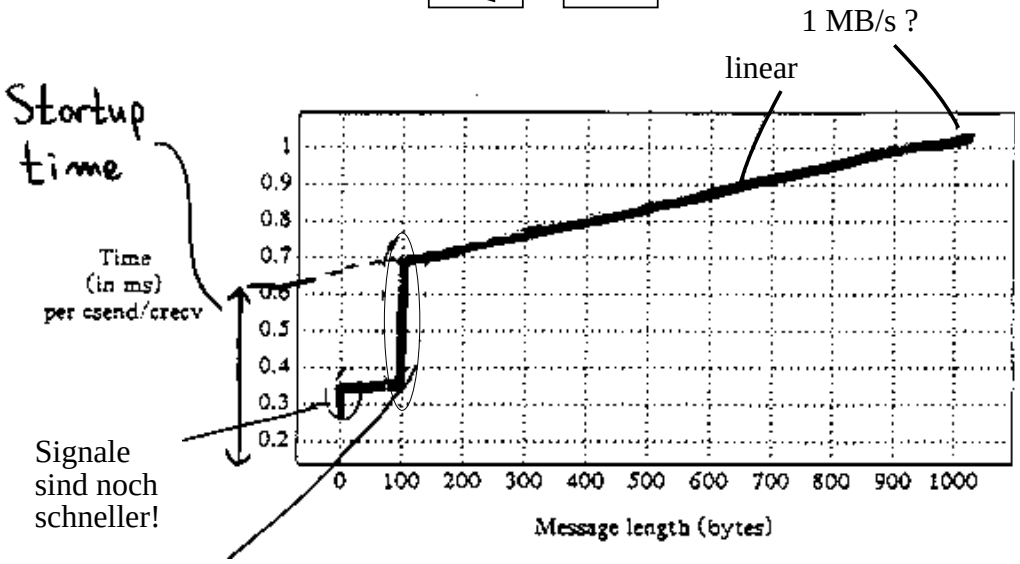
"Monster," I shrieked, "be thou juggler, enchanter, dream, or devil, no more will I endure thy mockeries. Either thou or I must perish." And saying these words I precipitated myself upon him.

Online-Text erhältlich bei: http://ebbs.english.vt.edu/20th/txts/abbott/guten.boiler.html oder: http://wiretap.spies.com/ftp.items/Library/Classic/flatland.txt

Das *Buch* ist erhältlich in mehreren Ausgaben; z.B.: Abbott, Edwin A.: Flatland. Penguin, 1987, ISBN: 0140076158, DM 11,90

Die Intel-Hypercube-Parallelrechner

- Forschungsprototyp “Cosmic Cube”: Caltech (vor 1985)
- iPSC/1 ca. ab 1985; iPSC/2 ca. ab 1988, dann iPSC/860
- Typische Grössen: 16, 32, 64, 128 Knoten
  - Teilwürfel sind unabhängig nutzbar
- Spätere Maschinen nutzen Gittertopologie statt Hypercube!
- Hier: Messung der Kommunikationsleistung
  - Technik: Round-trip-Zeit

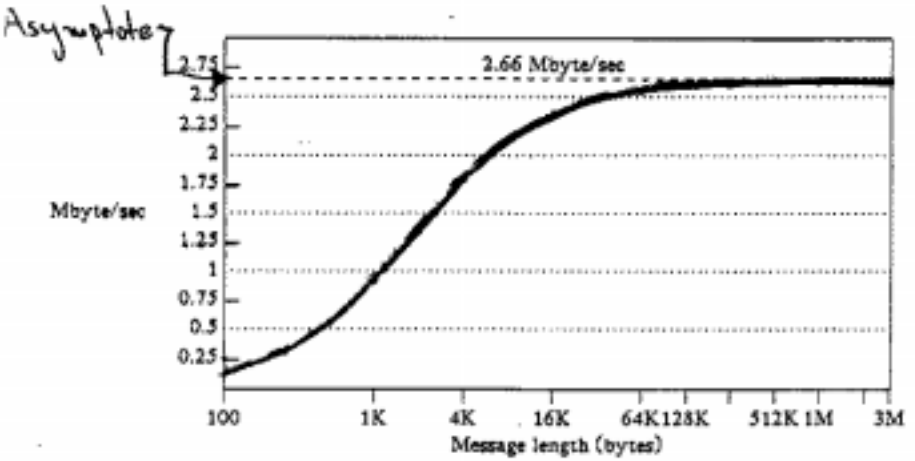


Bei Nachrichten > 100 Byte: zunächst überprüfen, ob der Empfänger genügend Pufferkapazität hat!

Man lese dazu folgenden Artikel: L. Bomans and D. Roose, *Benchmarking the iPSC/2 Hypercube Multiprocessor*, Concurrency - Practice and Experience, Vol.1 No 1, pp. 3-18, 1989

Asymptotische Übertragungsrate

- Konsequenz der relativ grossen Startup-Zeit beim iPSC: 2.8 MB/s “peak performance” ist ein asymptotischer Wert; er gilt nicht für “normale” Nachrichten!



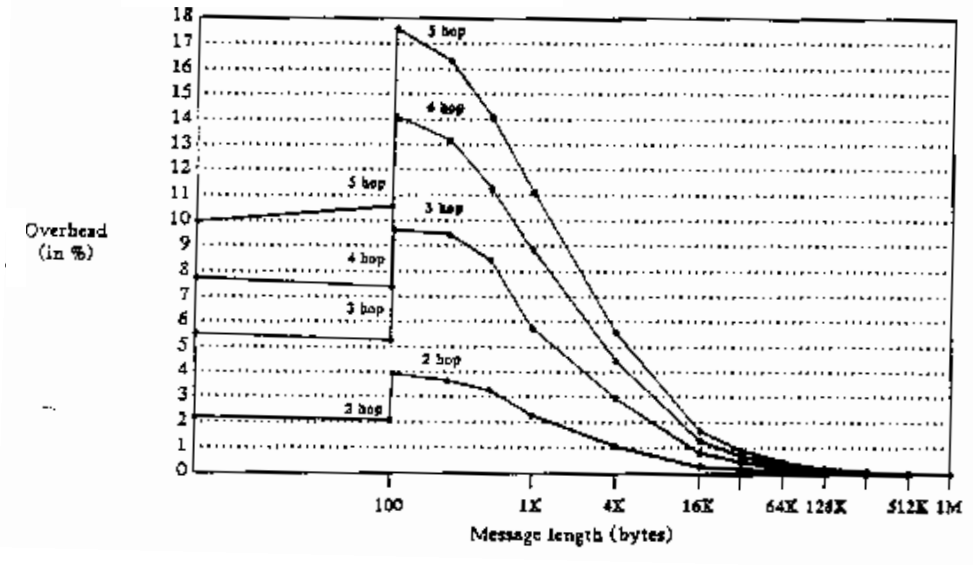
- Konstanter Overhead amortisiert sich erst bei sehr grossen Nachrichten

Multi-hop-Kommunikation

Kommunikation zwischen nicht benachbarten Knoten:
Wie schnell ist die Kommunikation dabei beim iPSC?

--> Zusatzaufwand fast vernachlässigbar!

iPSC/2-Hypercube der Dimension 6 --> maximal 5 “hops”

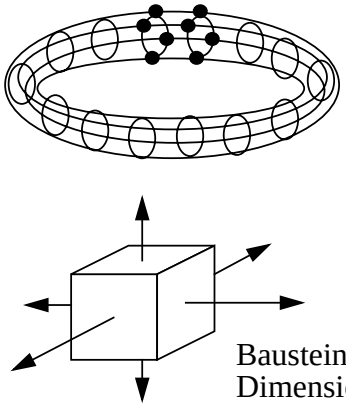


Weitere Frage:

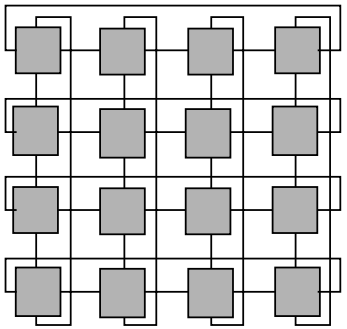
Gegenseitige Behinderung verschiedener Nachrichten wegen gemeinsam benutzter Verbindungen?
(“Channel contention”)

Eine andere Verbindungstopologie: der d-dimensionale Torus

= d-dimensionales “wrap-around Gitter”



Baustein für 3 Dimensionen



2 Dimensionen

- Rekursives Konstruktionsprinzip: „Nimm w_{d-1} gleiche Tori der Dimension d -1 und verbinde korrespondierende Elemente zu Ring“

- Bei Ausdehnung w_i in Dimension i:
  $n = w_1 \times w_1 \times .. \times w_d$ Knoten;
  mittlere Entfernung zw. 2 Knoten: $\Delta \approx \frac{1}{4} \sum w_i$

- *Ring* als Sonderfall d = 1 !

- *Hypercube* der Dimension d ist d-dimensionaler Torus mit $w_i = 2$ für alle Dimensionen!

  --> $\Delta = \frac{1}{4} \sum_d 2 = \frac{1}{4}(2\,d) = \frac{d}{2} = \underline{\underline{\frac{1}{2} \log_2 n}}$