Presenting Data from Experiments in Algorithmics

Restrictions

- black and white $\rightsquigarrow$ easy and cheap printing
- 2D (stay tuned)
- no animation
- no realism desired

Not here

- ensuring reproducibility
- describing the setup
- finding sources of measurement errors
- reducing measuremnt errors (averaging, median,unloaded machine. . .)
- measurements in the creative phase of experimental algorithmics.

The Starting Point

- (Several) Algorithm(s)
- A few quantities to be measured: time, space, solution quality, comparisons, cache faults,... There may also be measurement errors.
- An unlimited number of potential inputs. $\rightsquigarrow$ condense to a few characteristic ones (size, $|V|$, $|E|$, ... or problem instances from applications)

Usually there is not a lack but an abundance of data $\neq$ many other sciences

The Process

Waterfall model?

1. Design
2. Measurement
3. Interpretation

Perhaps the paper should at least look like that.

The Process

- Eventually stop asking questions (Advisors/Referees listen !)
- build measurement tools
- automate (re)measurements
- Choice of Experments driven by risk and opportunity
- Distinguish mode

  explorative: many different parameter settings, interactive, short turnaround times

  consolidating: many large instances, standardized measurement conditions, batch mode, many machines

Of Risks and Opportunities

Example: Hypothesis = my algorithm is the best

big risk: untried main competitor

small risk: tuning of a subroutine that takes 20 % of the time.

big opportunity: use algorithm for a new application

$\rightsquigarrow$ new input instances

Basic Principles

- Minimize nondata ink
  (form follows function, not a beauty contest,...)
- Letter size ≈ surrounding text
- Avoid clutter and overwhelming complexity
- Avoid boredom (too little data per m^2.
- Make the conclusions evident

Tables

+ easy
− easy $\leadsto$ overuse
+ accurate values ($\neq$ 3D)
+ more compact than bar chart
+ good for unrelated instances (e.g. solution quality)
− boring
− no visual processing

rule of thumb that “tables usually outperform a graph for small data sets of 20 numbers or less” [Tufte 83]

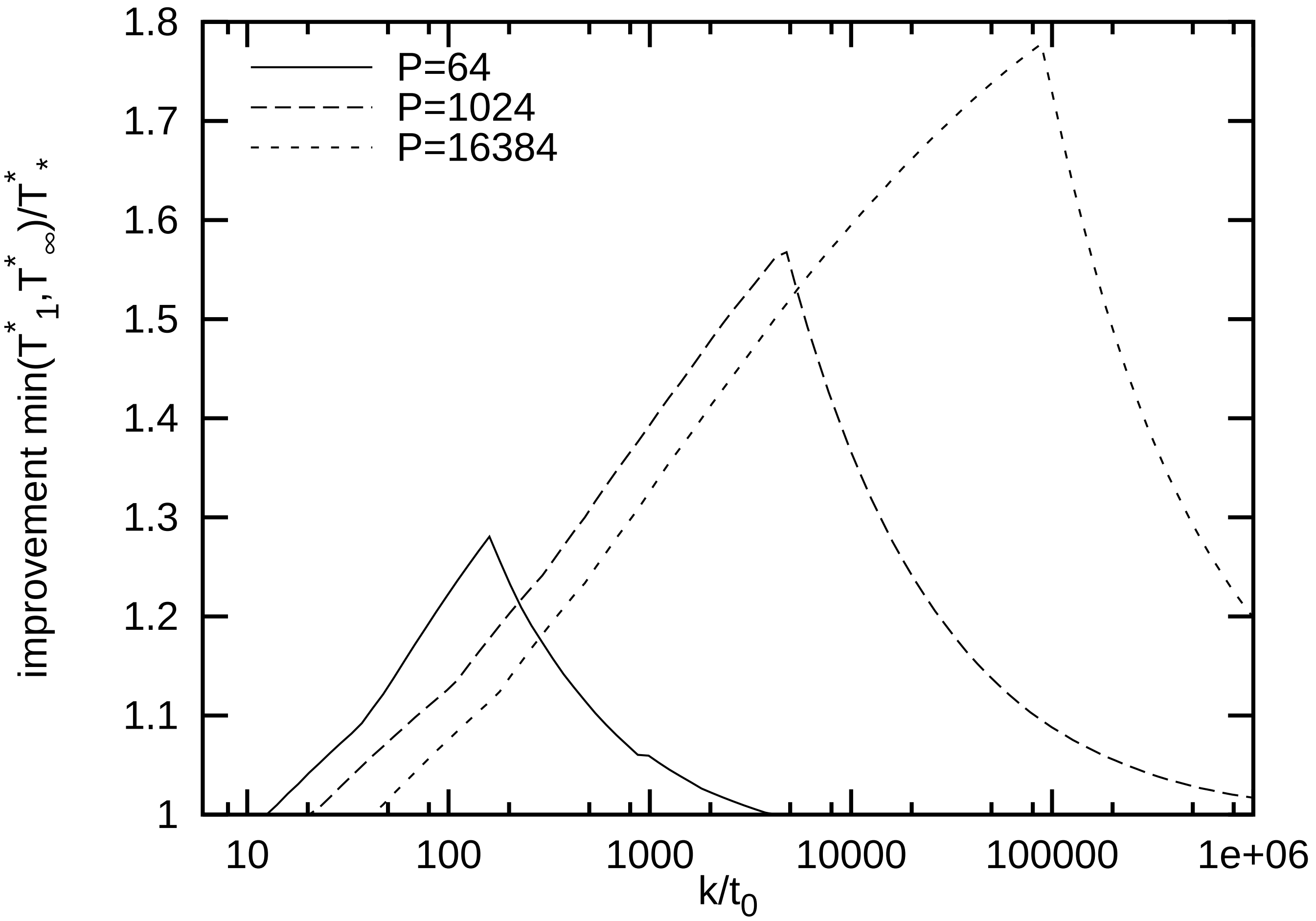
Curves in main paper, tables in appendix?

2D Figures

default: $x =$ input size, $y = f(\text{execution time})$

x Axis

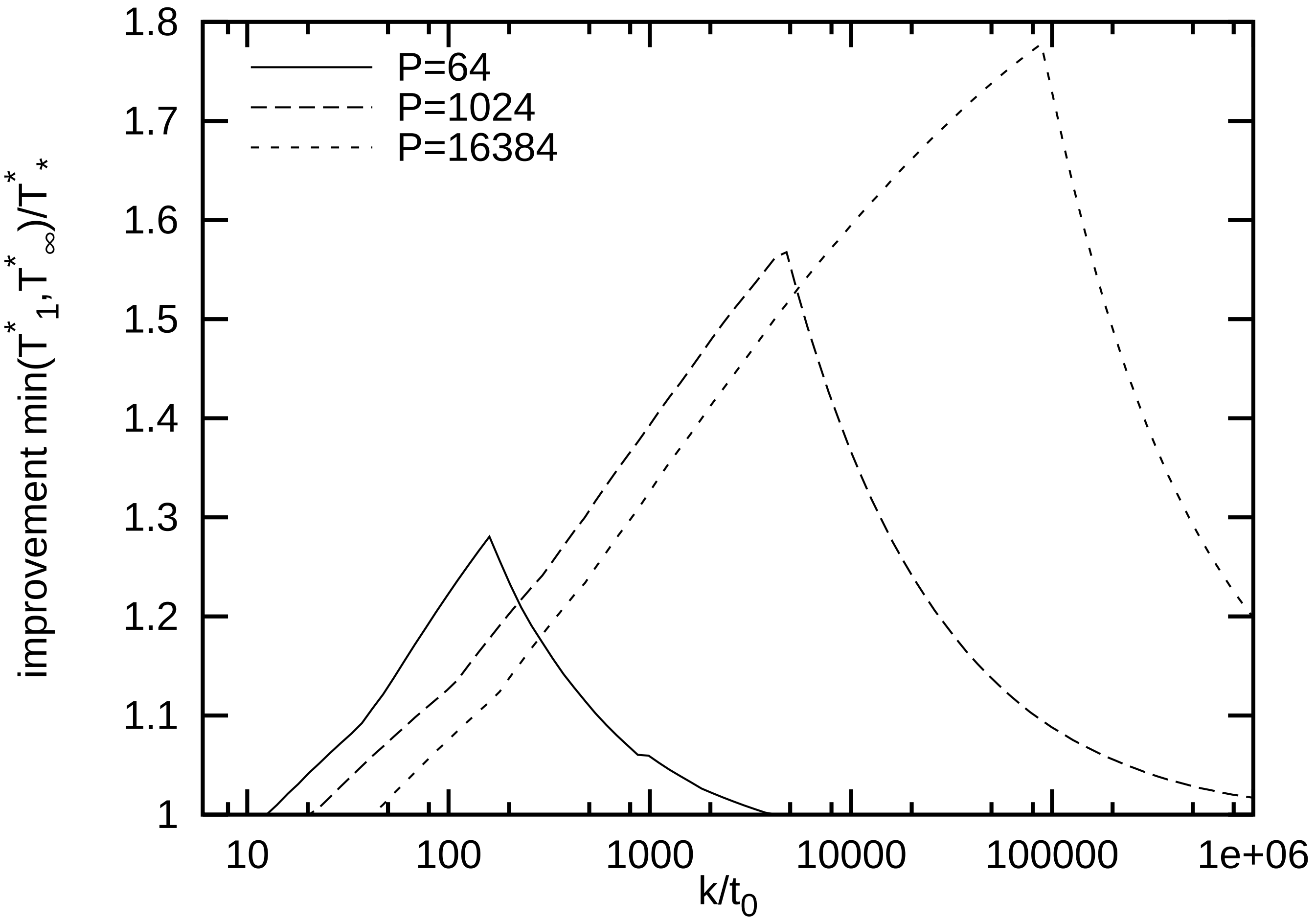
Choose unit to eliminate a parameter?



length k fractional tree broadcasting. latency $t_0 + k$

x Axis

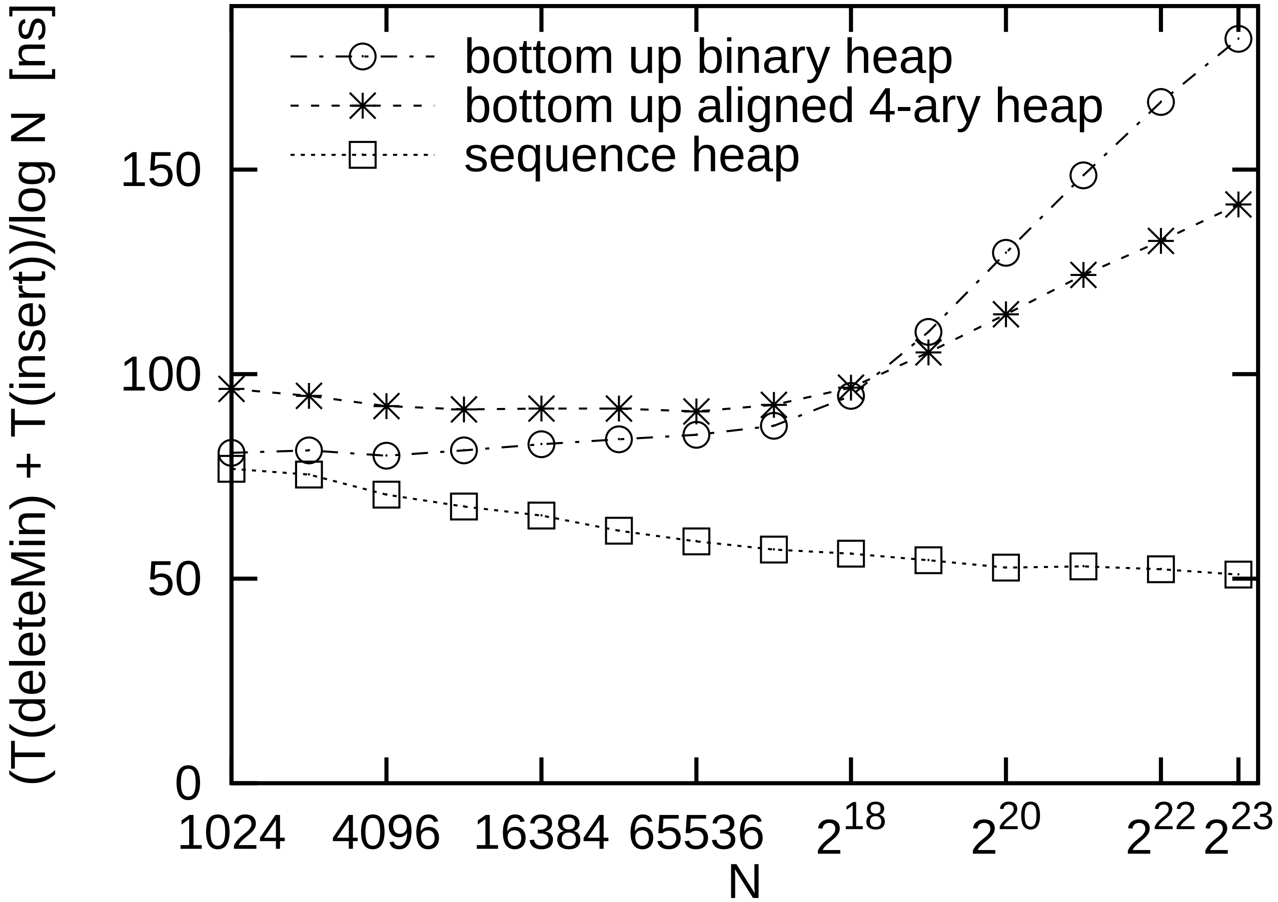
logarithmic scale?



yes if *x* range is wide

x Axis

logarithmic scale, powers of two (or $\sqrt{2}$)



with tic marks, (plus a few small ones)

gnuplot

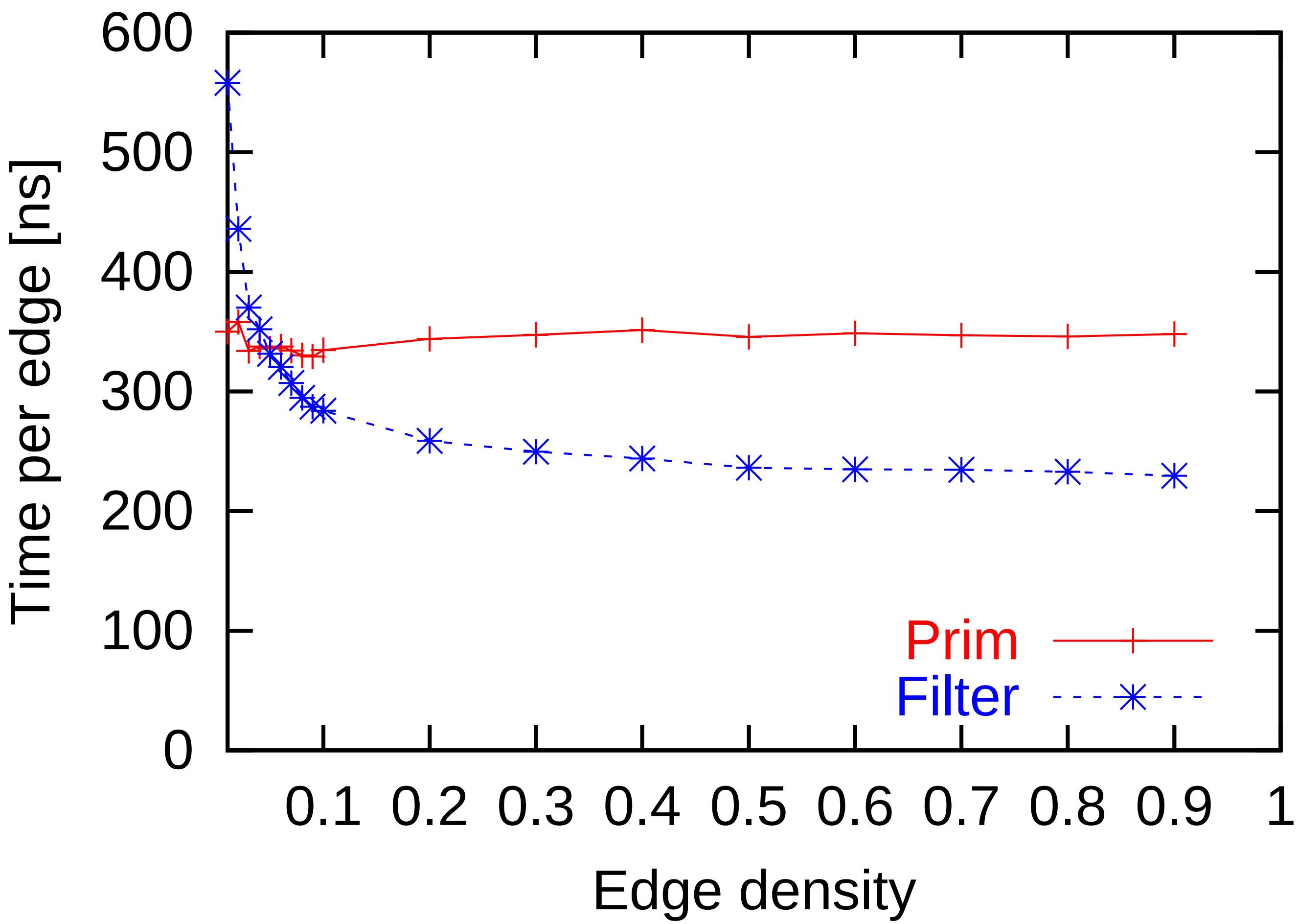
```
set xlabel "N"
set ylabel "(time per operation)/log N  [ns]"
set xtics (256, 1024, 4096, 16384, 65536, "2^{18}" 262144,
set size 0.66, 0.33
set logscale x 2
set data style linespoints
set key left
set terminal postscript portrait enhanced 10
set output "r10000timenew.eps"
plot [1024:10000000][0:220]\
 "h2r10000new.log" using 1:3 title "bottom up binary heap
 "h4r10000new.log" using 1:3 title "bottom up aligned 4-a
 "knr10000new.log" using 1:3 title "sequence heap" with l
```

Data File

```
256 703.125 87.8906
512 729.167 81.0185
1024 768.229 76.8229
2048 830.078 75.4616
4096 846.354 70.5295
8192 878.906 67.6082
16384 915.527 65.3948
32768 925.7 61.7133
65536 946.045 59.1278
131072 971.476 57.1457
262144 1009.62 56.0902
524288 1035.69 54.51
1048576 1055.08 52.7541
2097152 1113.73 53.0349
4194304 1150.29 52.2859
8388608 1172.62 50.9836
```

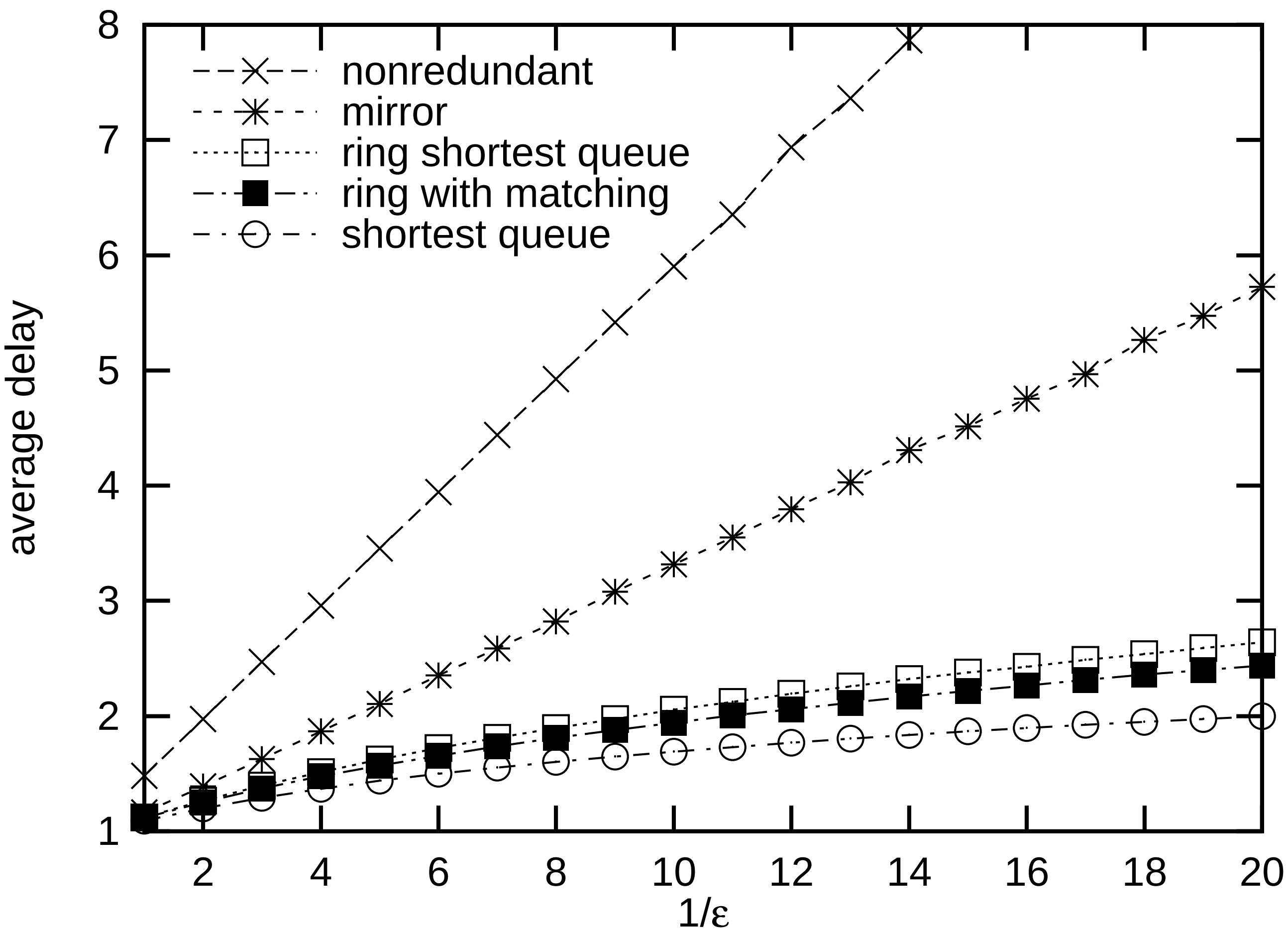
x Axis

linear scale for ratios or small ranges (#processor,. . .)



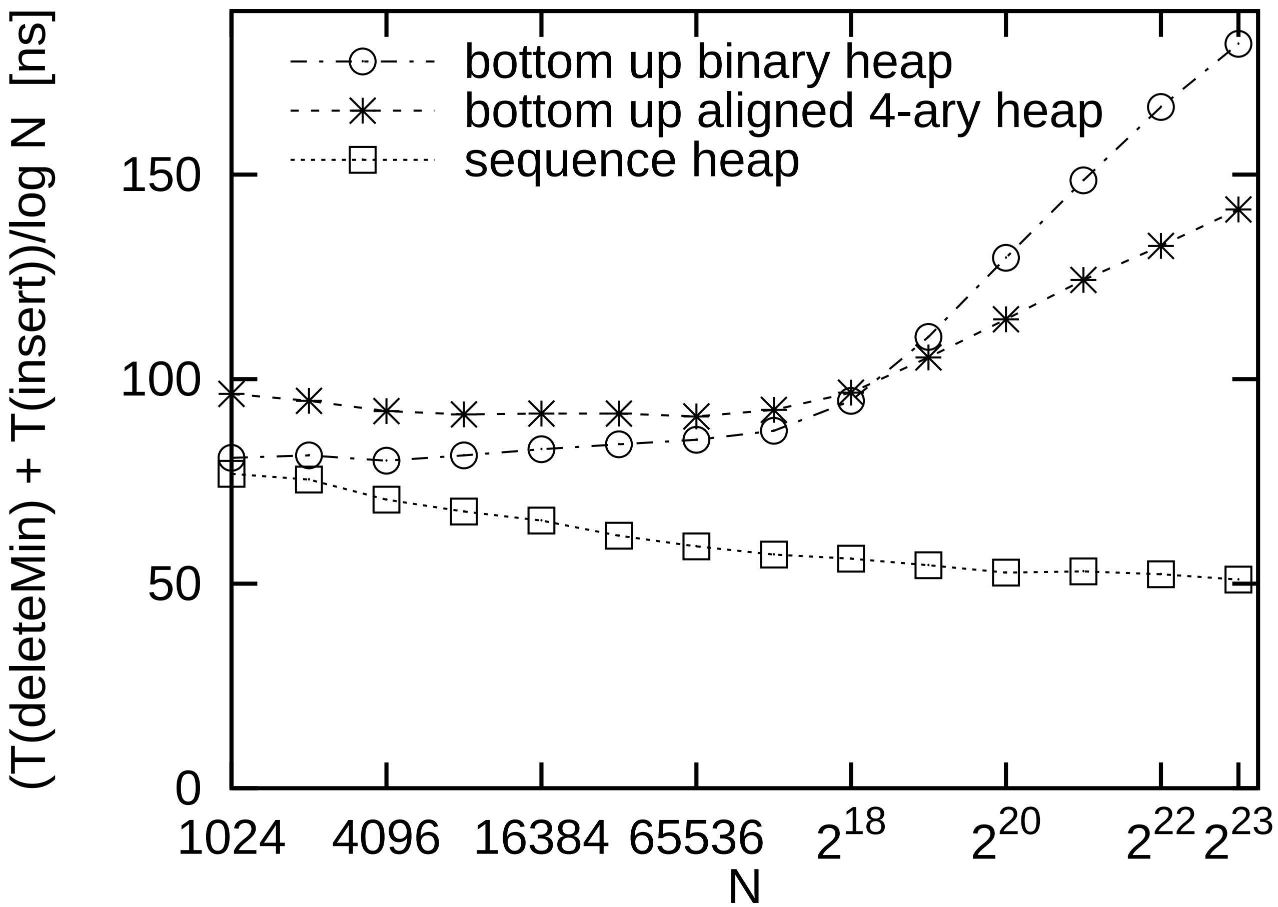
x Axis

An exotic scale: arrival rate 1 – ε of saturation point



y Axis

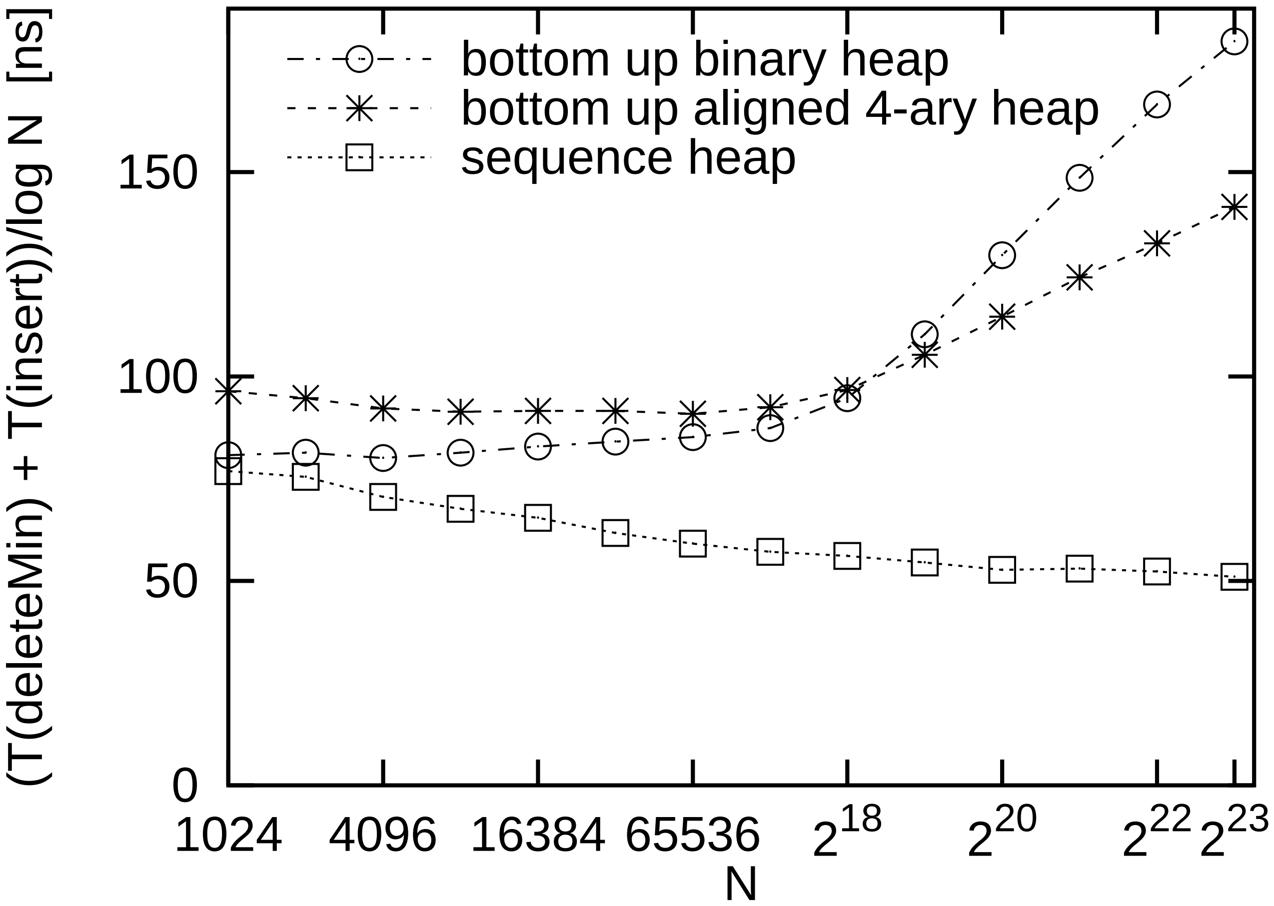
Avoid log scale ! scale such that theory gives ≈ horizontal lines



but give easy interpretation of the scaling function

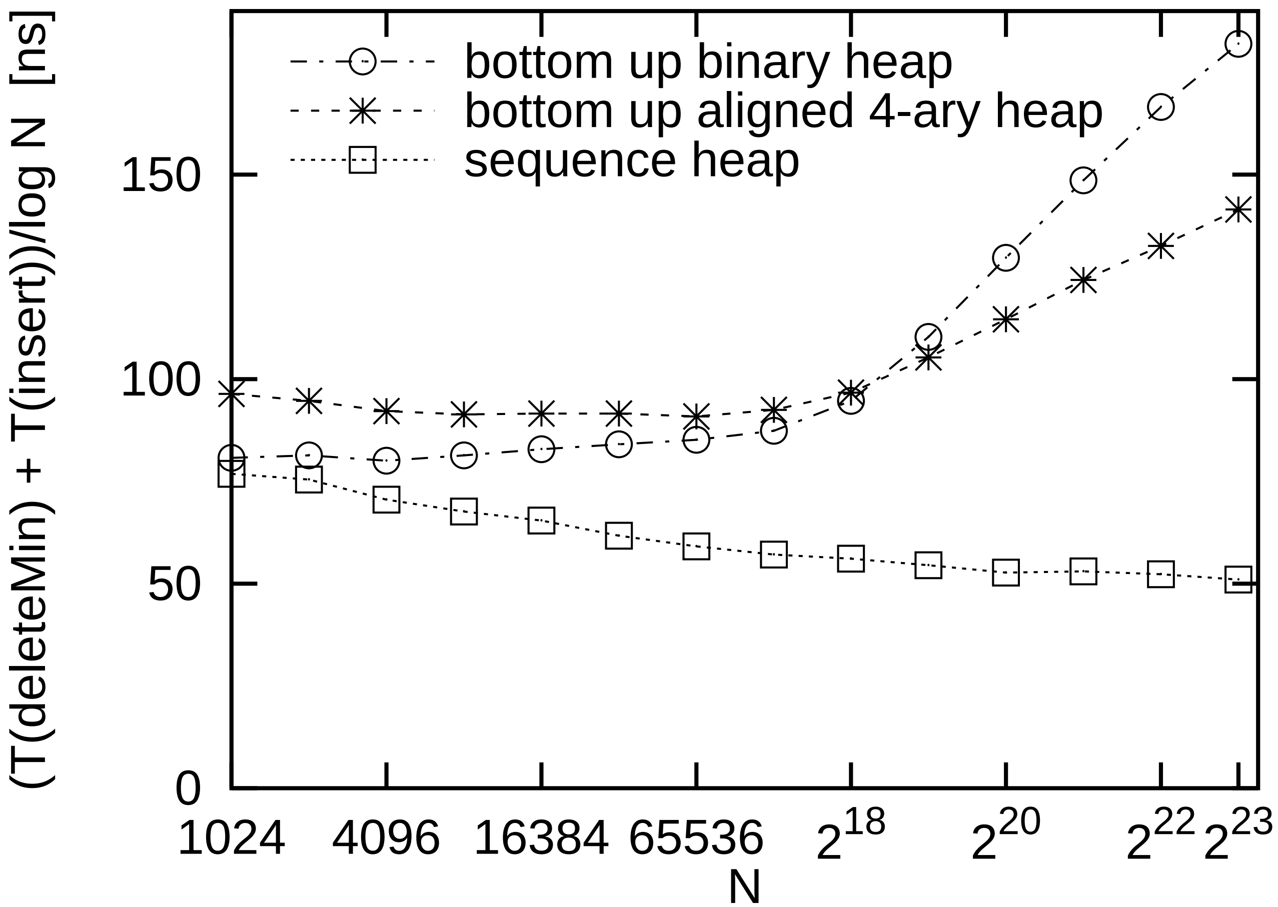
y Axis

give units



y Axis

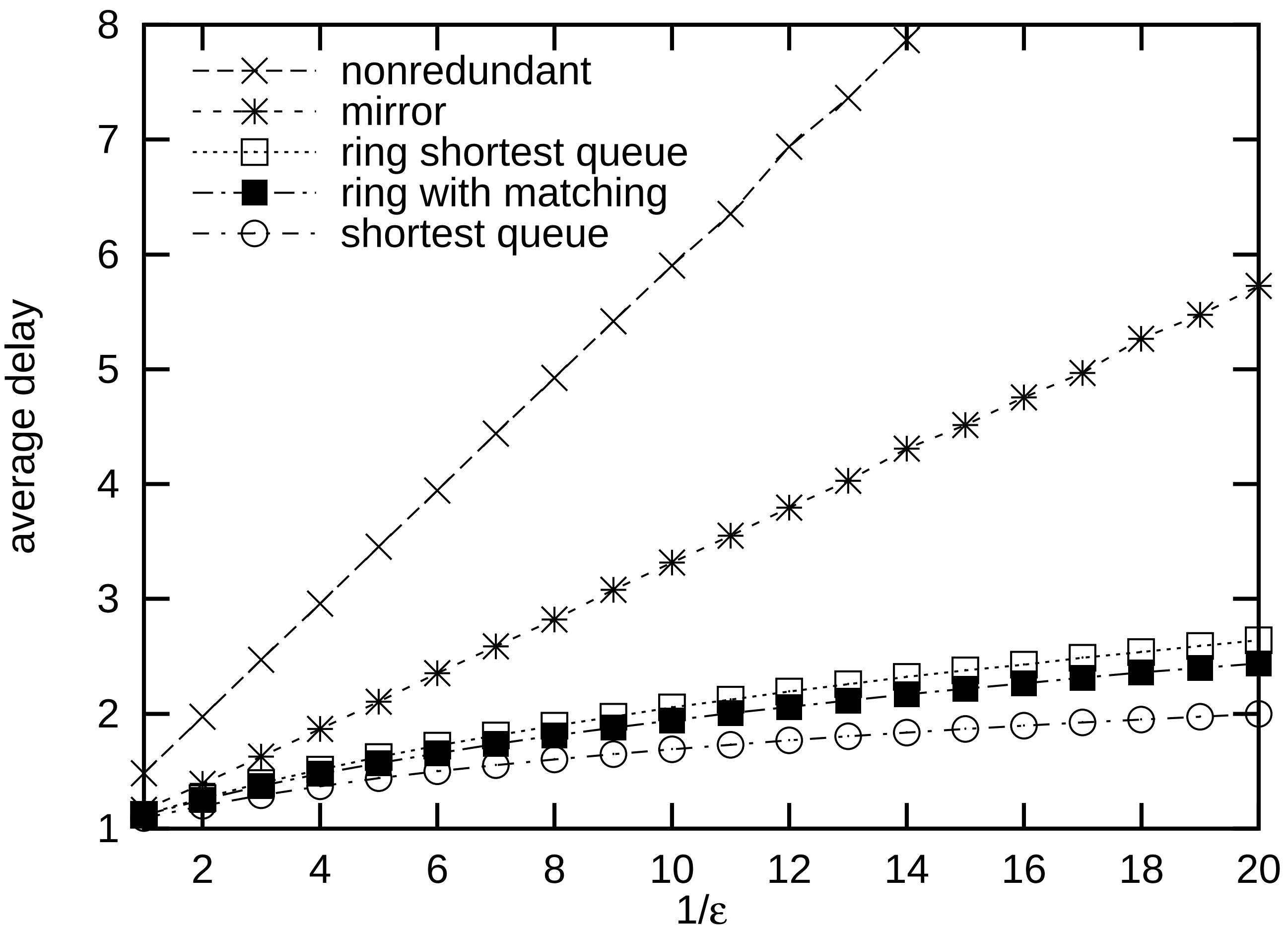
start from 0 if this does not waste too much space



you may assume readers to be out of Kindergarten

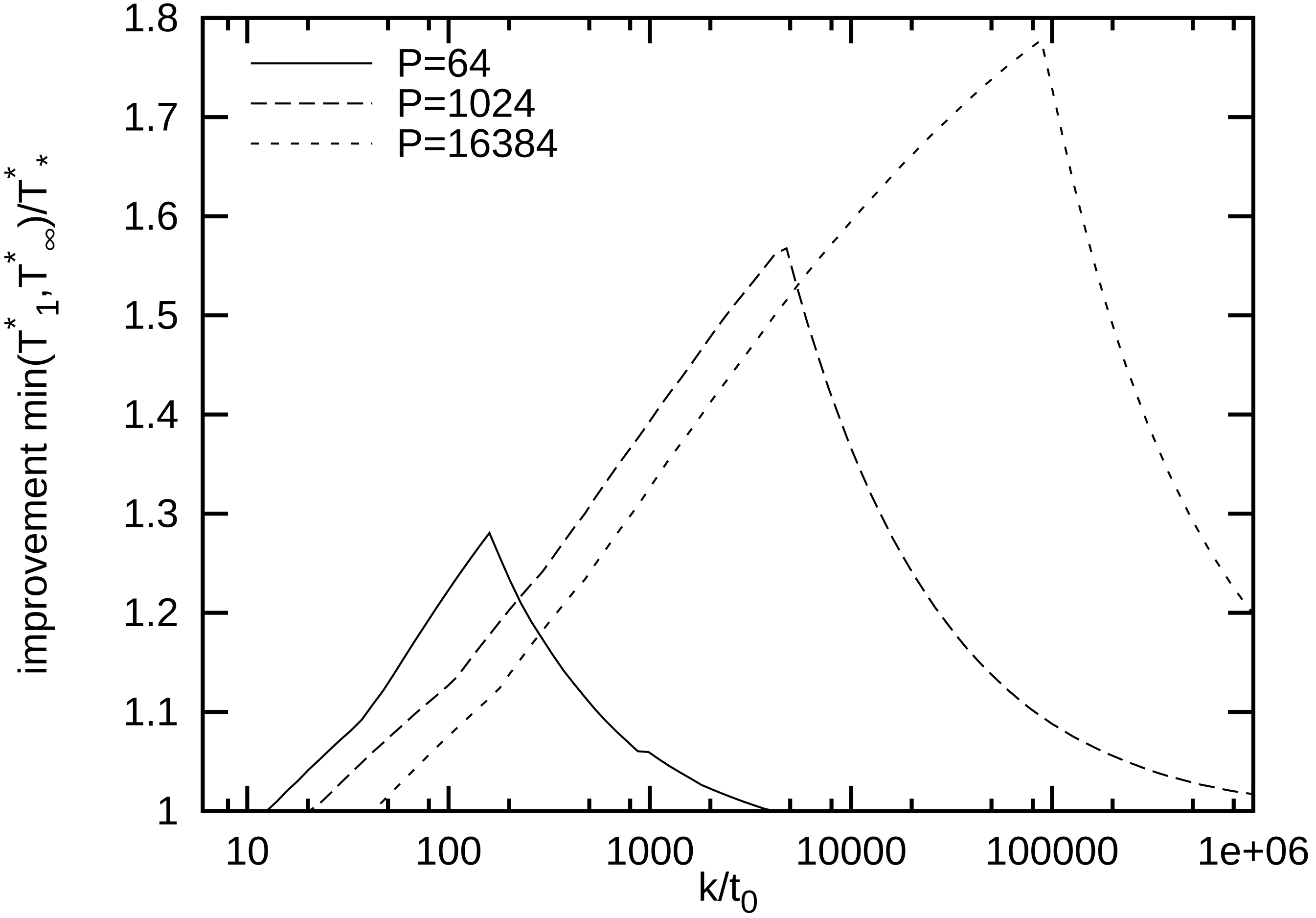
y Axis

clip outclassed algorithms



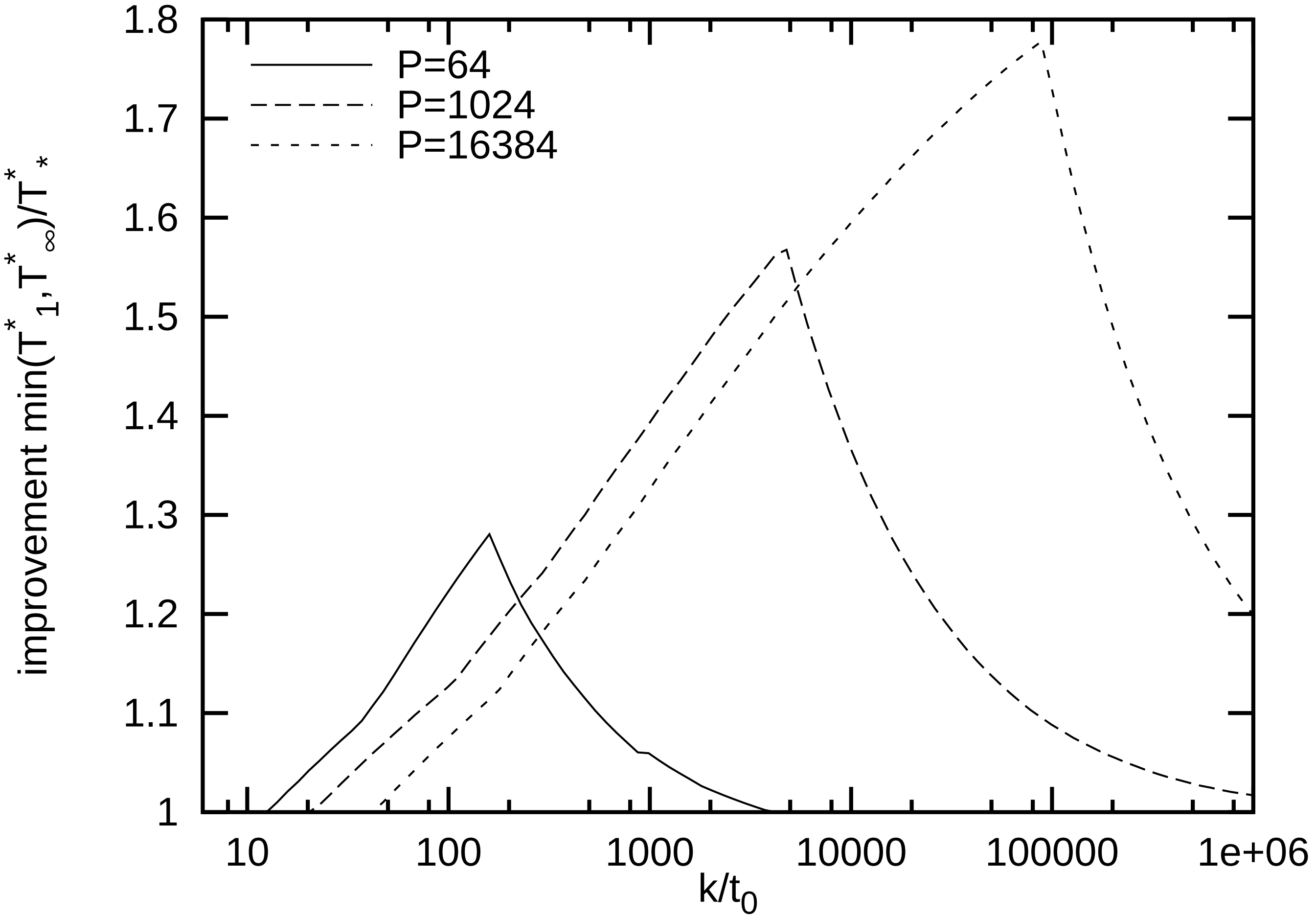
y Axis

vertical size: weighted average of the slants of the line segments in the figure should be about 45°[Cleveland 94]



y Axis

graph a bit wider than high, e.g., golden ratio [Tufte 83]



Multiple Curves

$+$ high information density

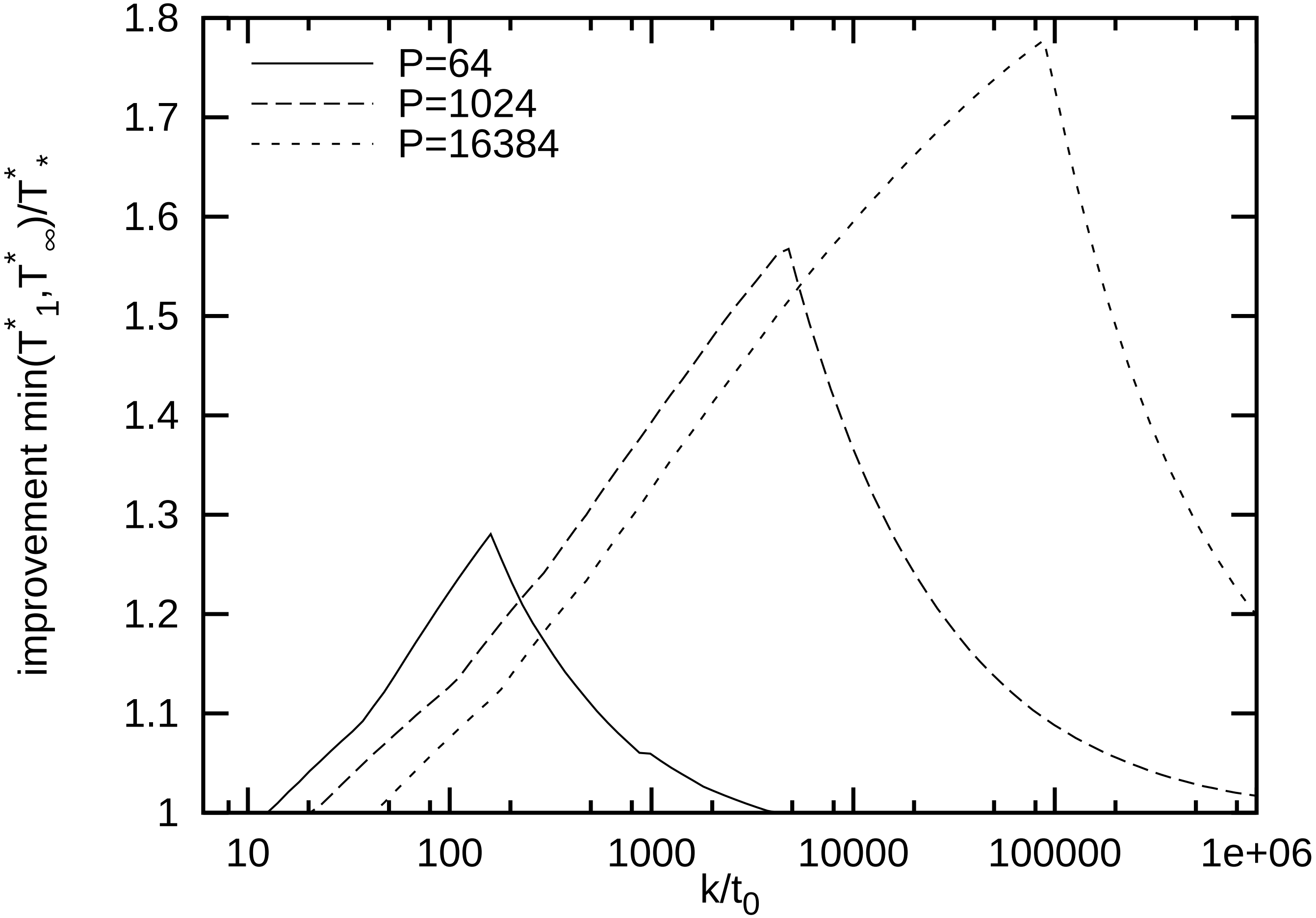
$+$ better than 3D (reading off values)

$-$ Easily overdone

≤ 7 smooth curves

Reducing the Number of Curves

use ratios



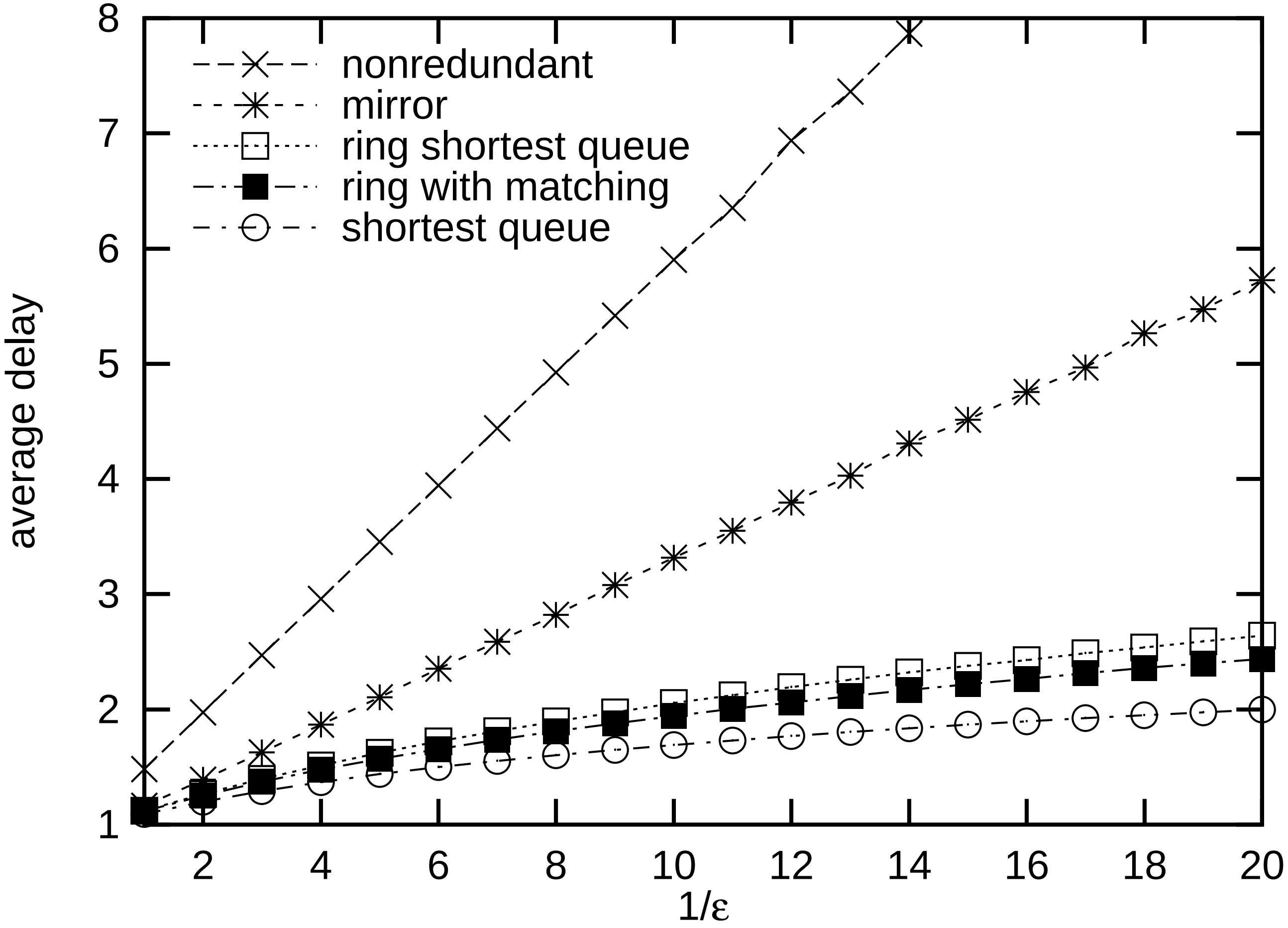
Reducing the Number of Curves

omit curves

- outclassed algorithms (for case shown)
- equivalent algorithms (for case shown)

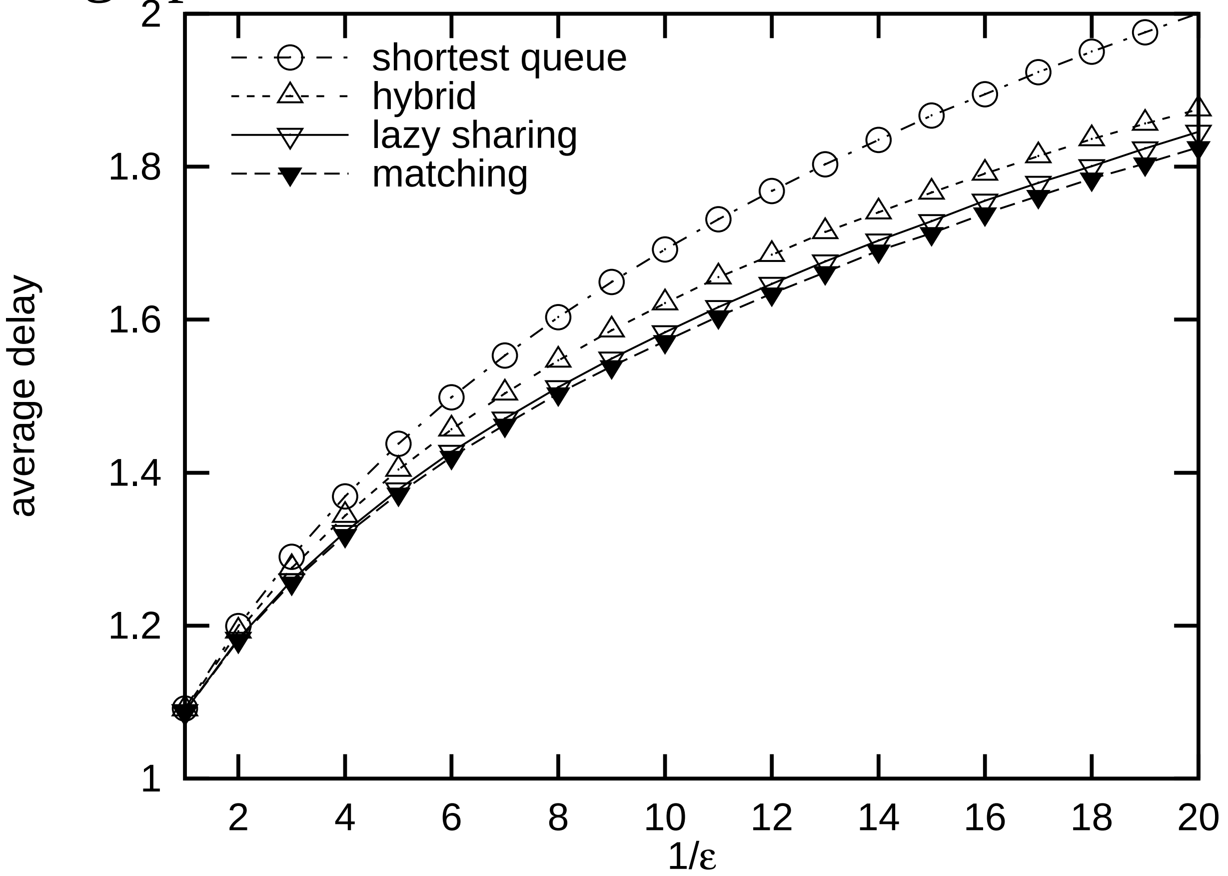
Reducing the Number of Curves

split into two graphs

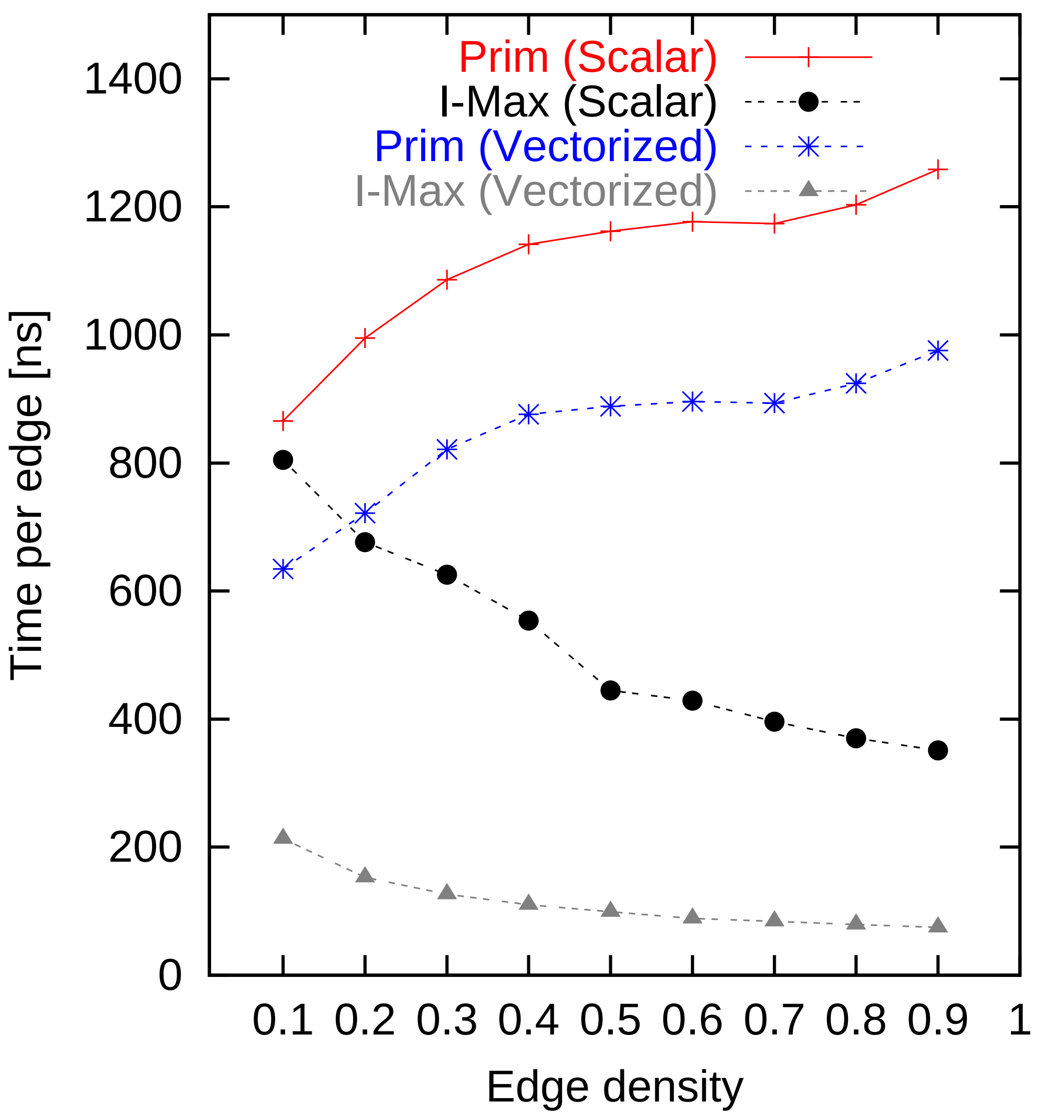


Reducing the Number of Curves

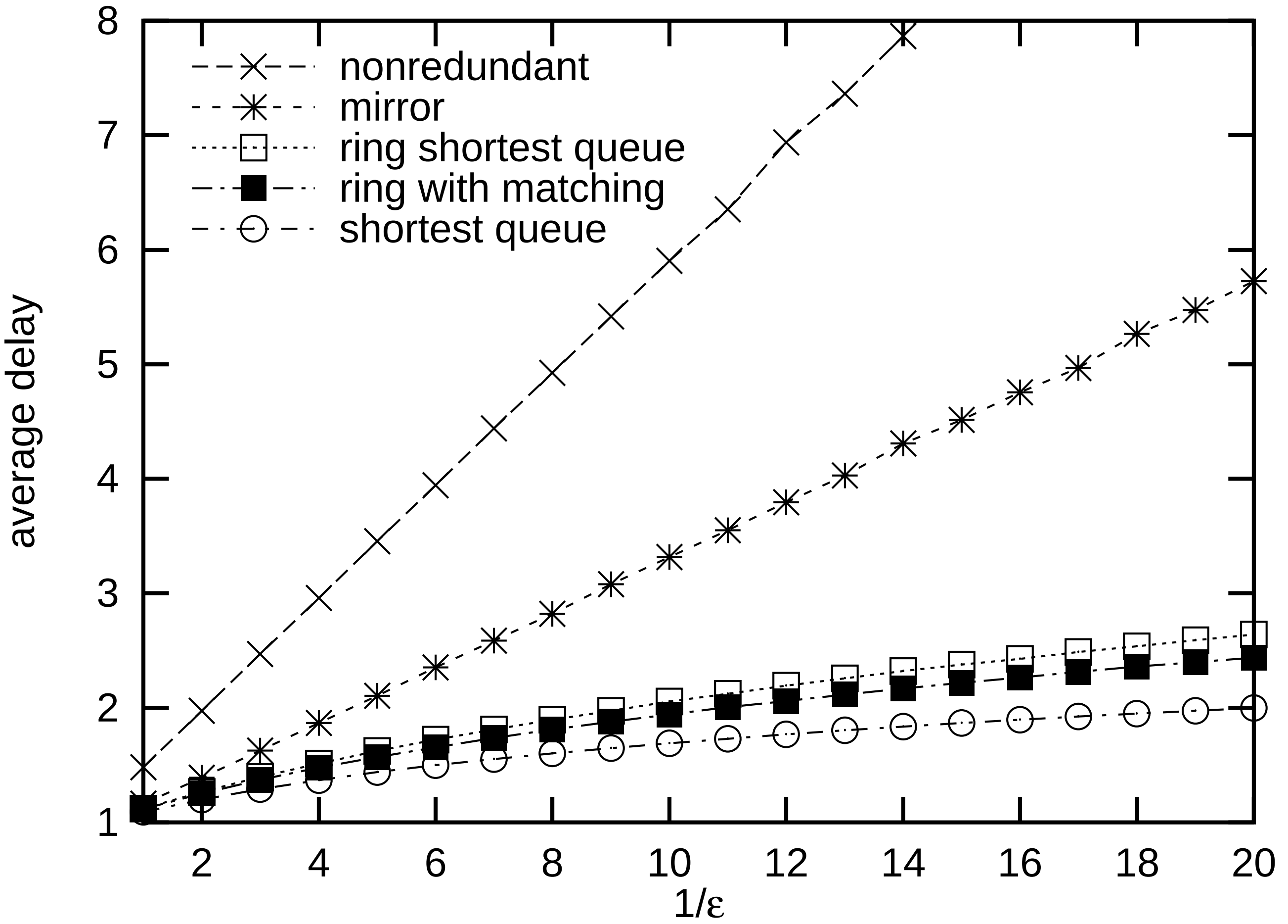
split into two graphs



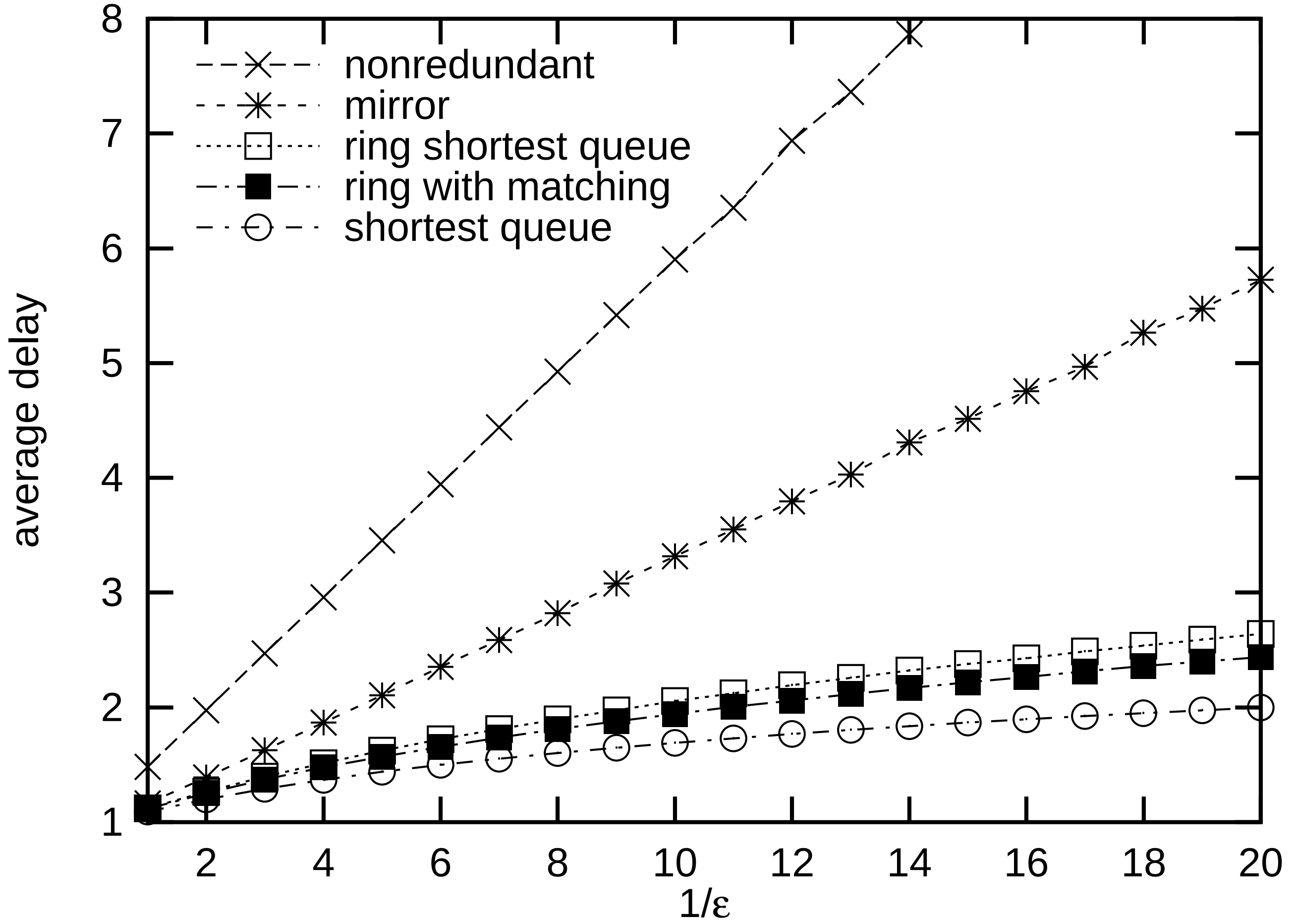
Keeping Curves apart: log y scale



Keeping Curves apart: smoothing



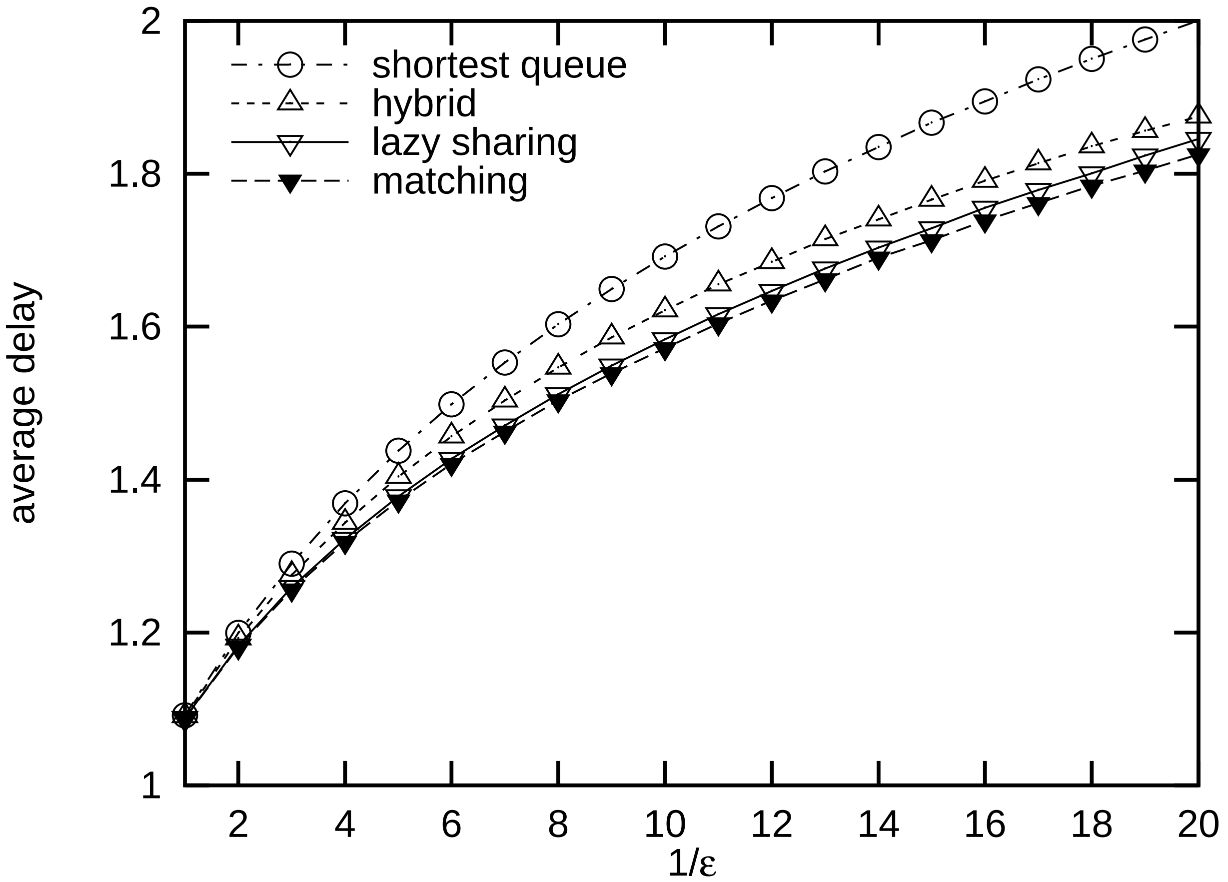
Keys



same order as curves

Keys

place in white space



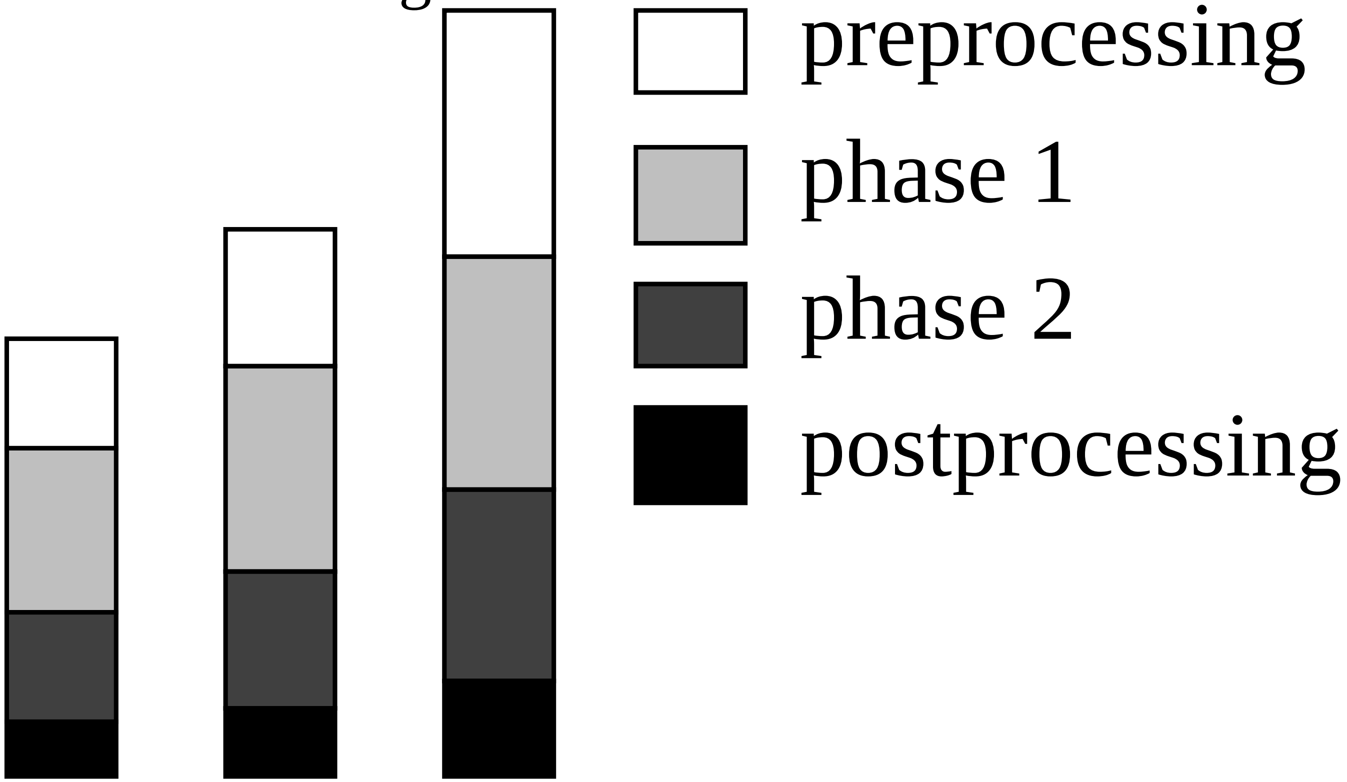
consistent in different figures

Todsünden

1. forget explaining the axes
2. connecting unrelated points by lines
3. mindless use/overinterpretation of double-log plot
4. cryptic abbreviations
5. microscopic lettering
6. excessive complexity
7. pie charts

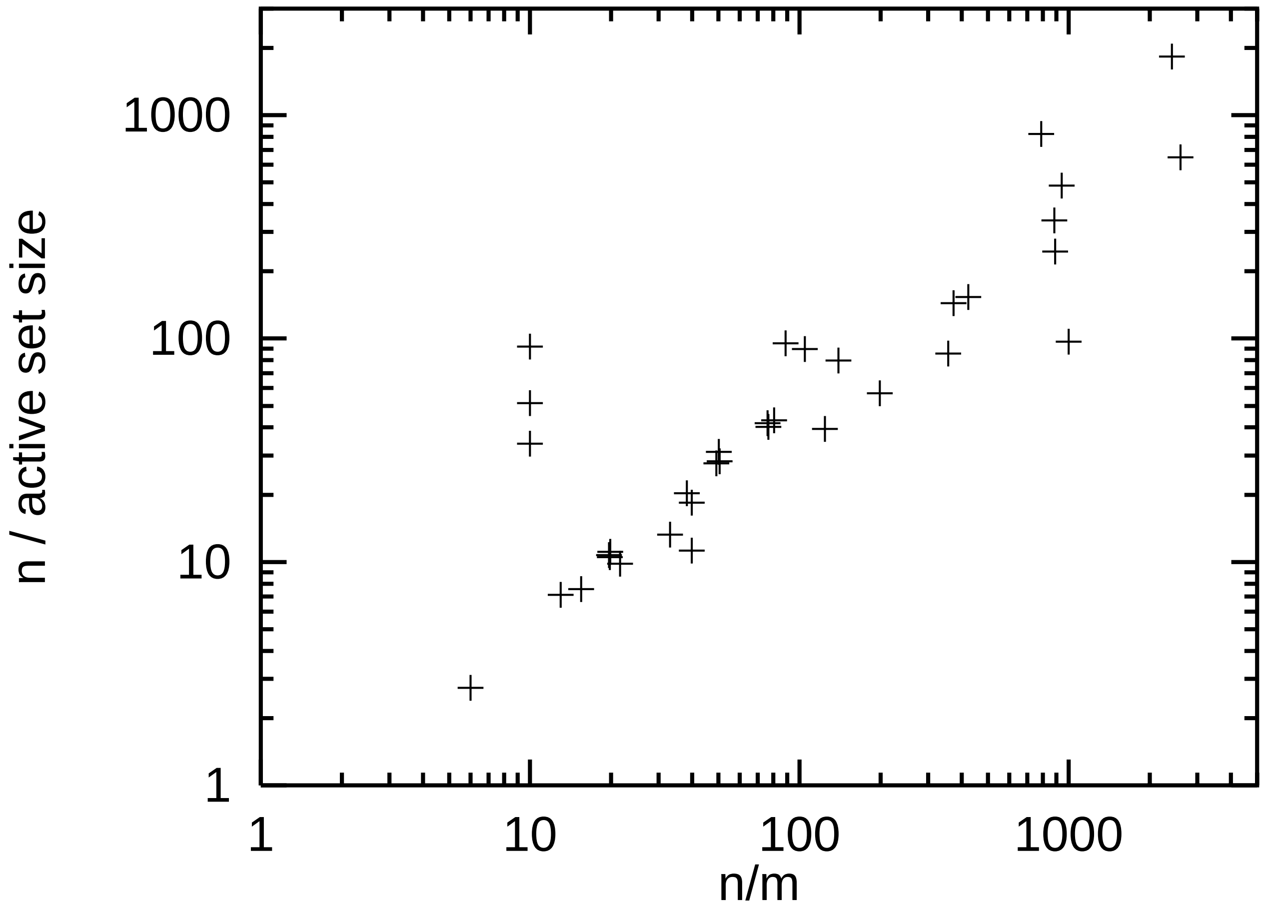
Arranging Instances

- bar charts
- stack components of execution time
- careful with shading



Arranging Instances

scatter plots



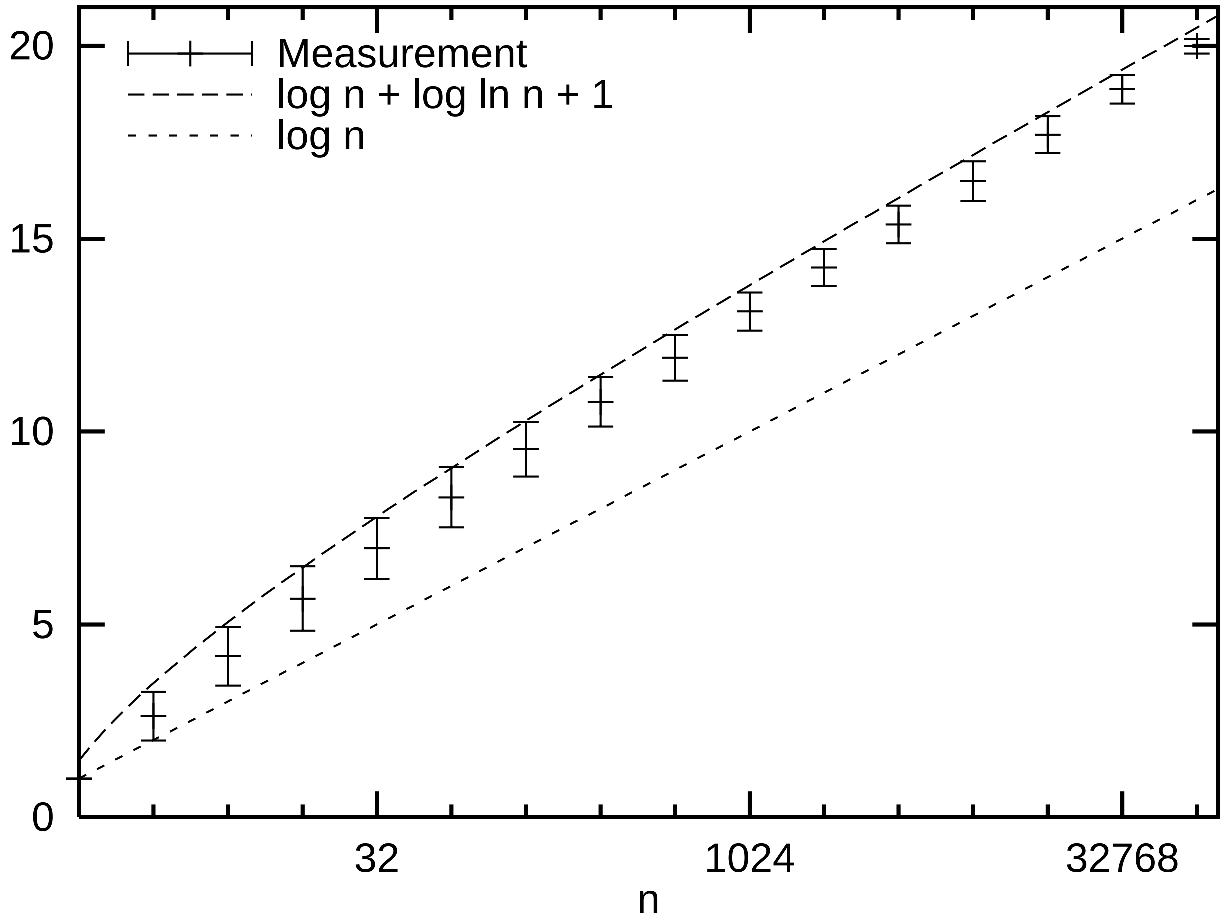
Measurements and Connections

- straight line between points do not imply claim of linear interpolation
- different with higher order curves
- no points imply an even stronger claim. Good for very dense smooth measurements.

Grids and Ticks

- Avoid grids or make it light gray
- usually round numbers for tic marks!
- sometimes plot important values on the axis

usually avoidable for randomized algorithms. median $\neq$ average,…



errors may not be of statistical nature!

3D

- – you cannot read off absolute values
- – interesting parts may be hidden
- – only one surface
- + good impression of shape

Caption

what is displayed

how has the date been obtained

surrounding text has more.

Check List

- Should the experimental setup from the exploratory phase be redesigned to increase conciseness or accuracy?
- What parameters should be varied? What variables should be measured? How are parameters chosen that cannot be varied?
- Can tables be converted into curves, bar charts, scatter plots or any other useful graphics?
- Should tables be added in an appendix or on a web page?
- Should a 3D-plot be replaced by collections of 2D-curves?
- Can we reduce the number of curves to be displayed?
- How many figures are needed?

- Scale the *x*-axis to make *y*-values independent of some parameters?
- Should the *x*-axis have a logarithmic scale? If so, do the *x*-values used for measuring have the same basis as the tick marks?
- Should the *x*-axis be transformed to magnify interesting subranges?
- Is the range of *x*-values adequate?
- Do we have measurements for the right *x*-values, i.e., nowhere too dense or too sparse?
- Should the *y*-axis be transformed to make the interesting part of the data more visible?
- Should the *y*-axis have a logarithmic scale?

- Is it be misleading to start the y-range at the smallest measured value?
- Clip the range of y-values to exclude useless parts of curves?
- Can we use banking to 45°?
- Are all curves sufficiently well separated?
- Can noise be reduced using more accurate measurements?
- Are error bars needed? If so, what should they indicate? Remember that measurement errors are usually *not* random variables.
- Use points to indicate for which x-values actual data is available.
- Connect points belonging to the same curve.

- Only use splines for connecting points if interpolation is sensible.
- Do not connect points belonging to unrelated problem instances.
- Use different point and line styles for different curves.
- Use the same styles for corresponding curves in different graphs.
- Place labels defining point and line styles in the right order and without concealing the curves.
- Captions should make figures self contained.
- Give enough information to make experiments reproducible.