Measurements

Performance metric

The basic characteristics of a computer system that we need to measure are

- A count of how many times an event occurs (e.g processor initiates an input/output request)
- The duration of some time interval (e.g how long each of these requests take)
- The size of some parameter (e.g number of bits transmitted and stored)

These values can be used directly as performance metrics.

See David Lilja, Measuring Computer Performance, 2000

Performance metric

Speedup of a system 2 w.r.t. system 1 is defined to be a value β such that

$$\beta = T_1 / T_2$$

where T_1 and T_2 are the performance metric being compared.
Then, system 2 is β times faster than system 1.

System	**Time (s)**	**Speedup wrt System 1**
1	480	1
2	360	1.33
3	540	0.89
4	210	2.29

Note: if $\beta < 1$, then we mention *slowdown*

Performance metric

Relative change of a system 2 w.r.t. system 1 is defined to be a value Δ such that

$$\Delta = (T_1 - T_2)/\, T_2 = \beta - 1$$

where T_1 and T_2 are performance metrics being compared. Then, system 2 is $100 \cdot \Delta$ % faster than system 1.

System	**Time (s)**	**Speedup**	**RC wrt System 1**
1	480	1	0 %
2	360	1.33	33%
3	540	0.89	-11%
4	210	2.29	129%

Speed metric

Often, you are interested in normalizing event counts to a common time basis to provide a *speed metric*:

- Rate metric or throughput: divide the count of the number of events that occur in a given interval by the time interval over which the events occur

Typical example in communication networks:

- Network throughput: the rate of successful message delivery over a communication channel (e.g bits/s)

Speed metric

Speedup of a system 2 w.r.t. system 1 is defined to be a value β such that

$$\beta = R_2 / R_1$$

where R_1 and R_2 are "speed metrics" being compared. Then, system 2 is β times faster than system 1.

System	**bits/s**	**Speedup wrt System 1**
1	24	1
2	35	1.46
3	12	0.5
4	59	2.46

Speed metric

- Relative change of a system 2 w.r.t. system 1 is defined to be a value Δ such that

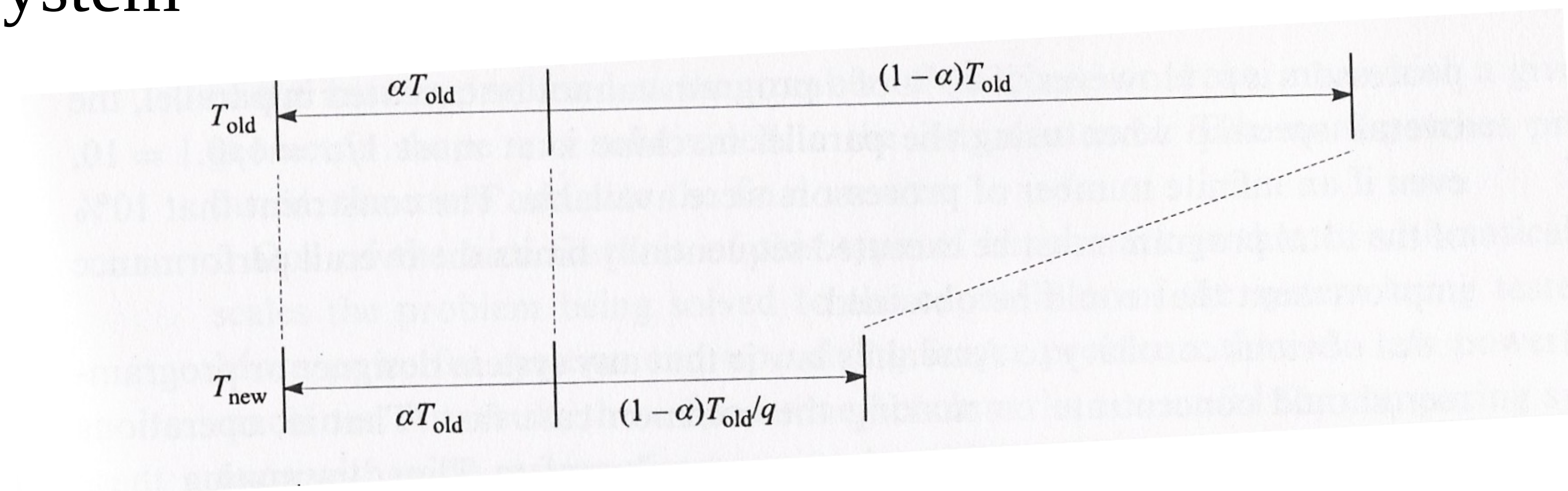
$$\Delta = (R_2 - R_1) / R_1 = \beta - 1$$

where R_1 and R_2 are "speed metrics" being compared. Then, system 2 is $100 \cdot \Delta$ % faster than system 1.

System	bits/s	Speedup	RC(%) wrt System 1
1	24	1	0%
2	35	1.46	46%
3	12	0.5	-50%
4	59	2.46	146%

Amdahl's law

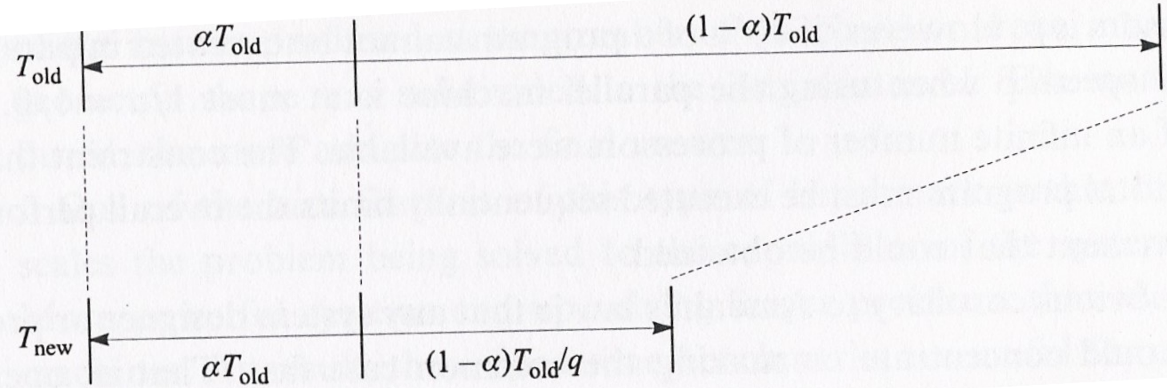
The overall performance improvement in a program is limited by that portion that is unaffected by whatever change made to the system



T_{old} and T_{new} is the time taken by the program before and after the improvement, respectively; α is the fraction of all operations that are unaffected by the improvement; q is the factor of improvement on the remaining operations.

Amdahl's law

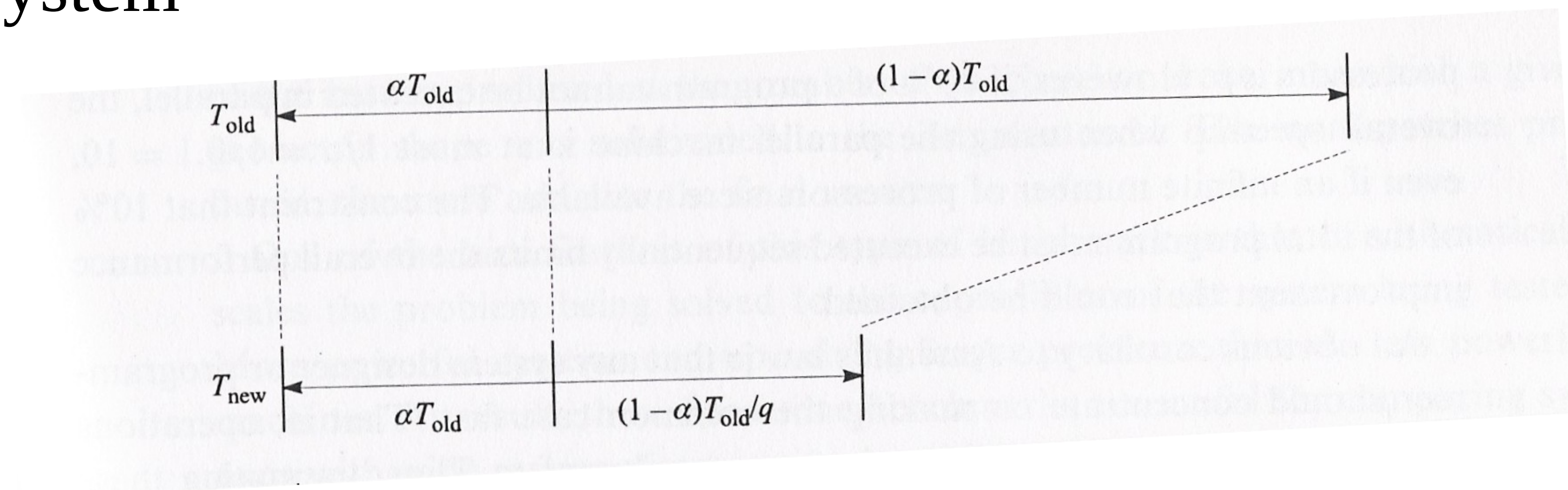
The overall performance improvement in a program is limited by that portion that is unaffected by whatever change made to the system



$$\beta = \frac{T_{old}}{T_{new}} = \frac{T_{old}}{\alpha T_{old} + (1-\alpha)T_{old}/q} = \frac{1}{1/q + \alpha(1-1/q)}$$

Amdahl's law

The overall performance improvement in a program is limited by that portion that is unaffected by whatever change made to the system



$$\lim_{q\to\infty} \beta = \lim_{q\to\infty} \frac{1}{1/q + \alpha(1\text{-}1/q)} = \frac{1}{\alpha}$$

Amdahl's law

The overall performance improvement in a program is limited by that portion that is unaffected by whatever change made to the system

$$\lim_{q\to\infty} \beta = \lim_{q\to\infty} \frac{1}{1/q + \alpha(1\text{-}1/q)} = \frac{1}{\alpha}$$

Example: If 10% of a program cannot be executed in parallel, the speedup for any *p* processors is at most $1/\alpha = 1/0.1 = 10$.

Properties of metrics

- Linearity
- Reliability
- Repeatability
- Consistency
- Independence

Properties of metrics

Linearity: If the value of the metric changes by a certain ratio, the actual performance of the machine should change by the same ratio.

Example: Suppose that you are upgrading your system to a new system whose speed metric is twice as large as the speed metric on your current system. Then, you expect the new system to able to run your application programs in half the time taken by your current system.
Intuitively appealing but not true for some metrics, e.g dB.

Properties of metrics

Reliability: A performance metric is reliable if a system A always outperforms system B when the corresponding values of the metric for both systems indicate that system A should outperforms system B.

Example: Suppose that you have a new metric called WIPS to compare the performance of two computers. WIPS on system A and B is 128 and 96 on word processing activities, respectively. WIPS is reliable if systems A outperforms system B when executing those applications.

Many computer performance metrics are unreliable, e.g MIPS.

Properties of metrics

Repeatability: A performance metric is repeatable if the same value of the metric is measured each time the same experiment is performed.

This implies that that a good metric is deterministic, which is usually a strong requirement.

Properties of metrics

Consistency: A consistent performance metric is one for which the units of the metric and its precise definition are the same across different systems.
Not all computer performance metrics are consistent, e.g MIPS and MFLOPS.

Independence: To avoid corruption of its meaning, a good metric should be independent of outside influence, e.g. computer manufacturers.
Not true also for computer performance metrics, see SPEC benchmarks.

Properties of metrics

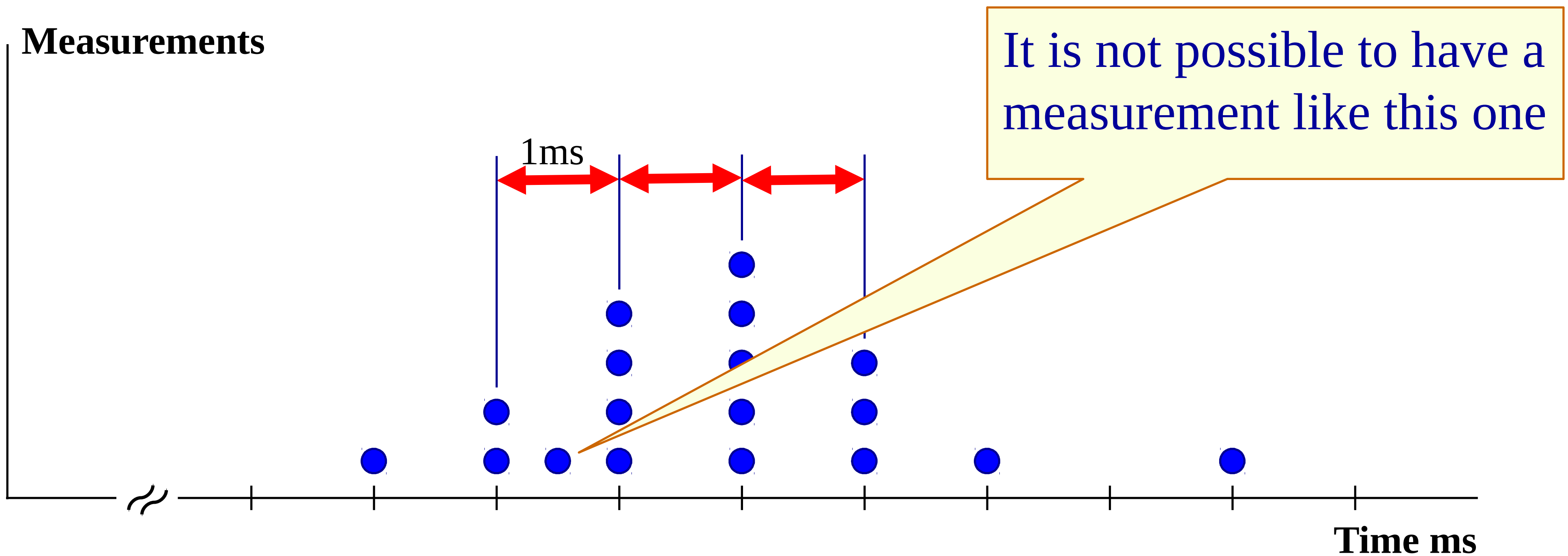
Homework: Investigate computer performance metrics with respect to these properties

- Clock rate
- MIPS (millions of instructions executed per second)
- MFLOPS (millions of floating-point operations executed per second)
- SPEC (System Performance Evaluation Cooperative)
- Execution time of a given application program (CPU vs Wall-clock time)

Resolution

Resolution of the measuring instrument: the smallest difference between measurements provided by a measuring device

Example: measuring execution time of a program in milliseconds



(adapted from Dror Feitelson, HU, slides)

Uncertainty

Uncertainty of the measurement: if we repeat a measurement we will get slightly different results. Reflects the lack of **precision** of the measurement

Two types of uncertainties (leading to errors):

- **Random uncertainties**

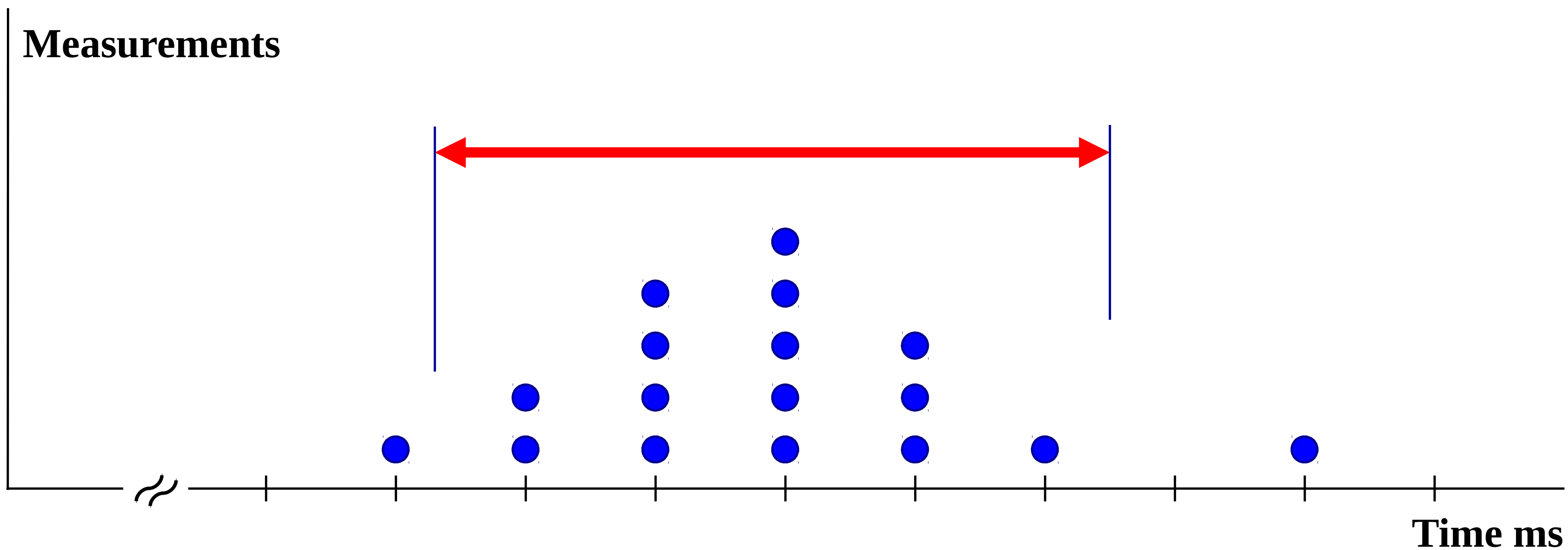
Variations in the measurements that occur without a predictable pattern.

- **Systematic uncertainties**

Variations that consistently cause the measured value to be smaller or larger that the exact value.

Random uncertainties

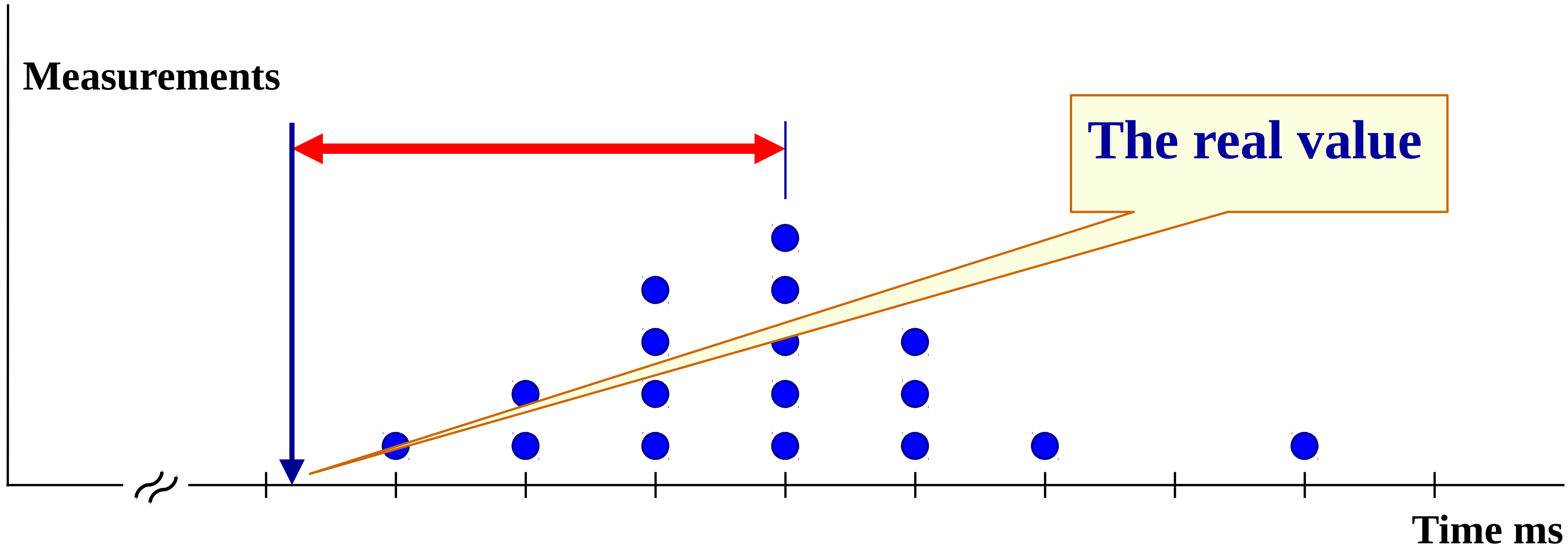
- Occur without a predictable pattern.
- Can be reduced but never eliminated.
- Must be statistically analyzed and reported in the measurement process.



(adapted from Dror Feitelson, HU, slides)

Systematic uncertainties

- Systematic deviations from the real value
- Due to many possible causes (e.g., inaccurate measuring tools, misscalibration, tool reaction time/delay, warm-up, etc)
- Once identified, can be eliminated (one of the steps in experiment design)



(adapted from Dror Feitelson, HU, slides)

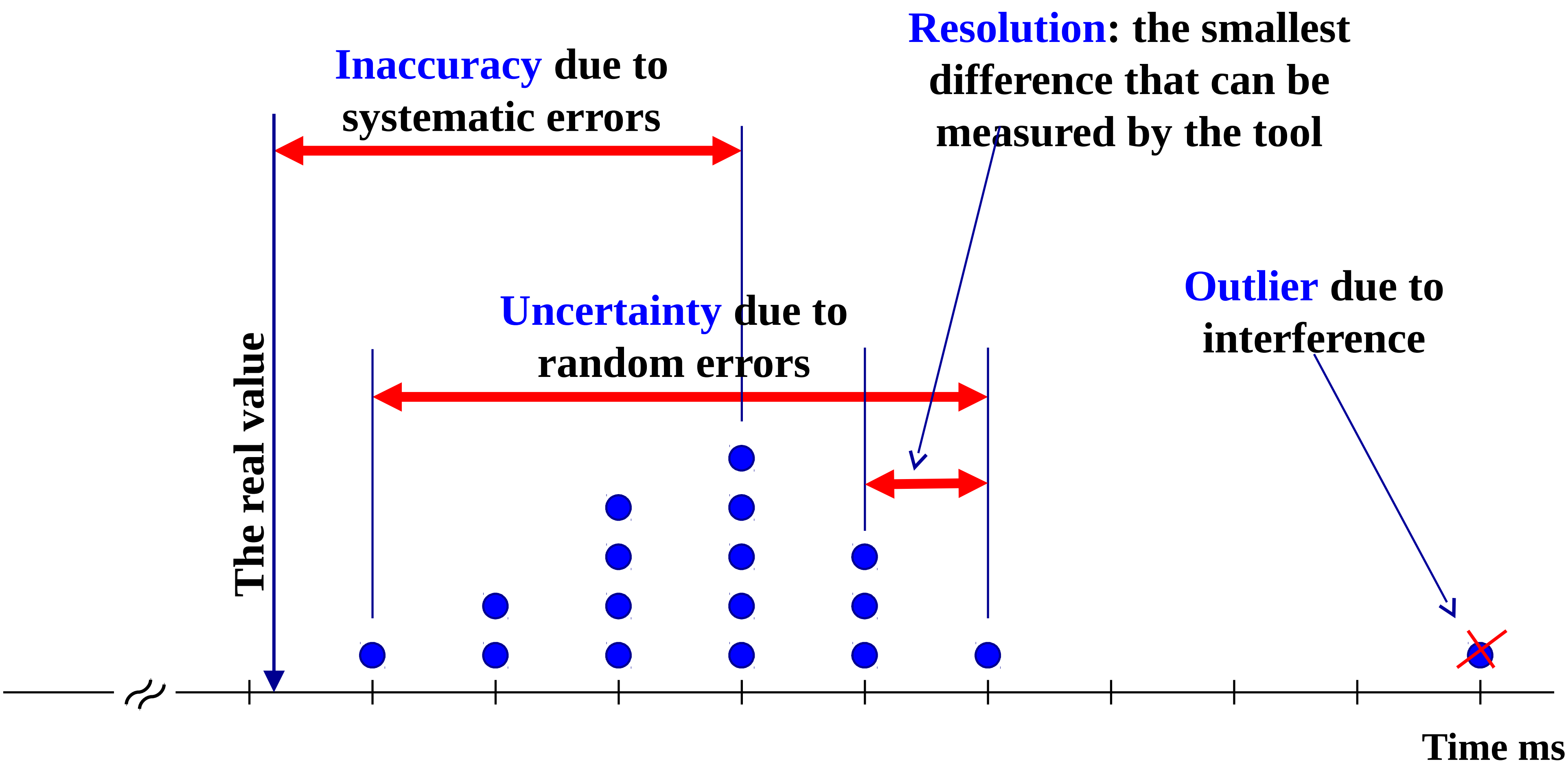
Uncertainty and variability

The variability in the measurements can result from two different sources:

- **Precision limitations of the measuring instrument.**
  - Even if the experiment conditions were totally stable, the different measurements would show slightly different values.

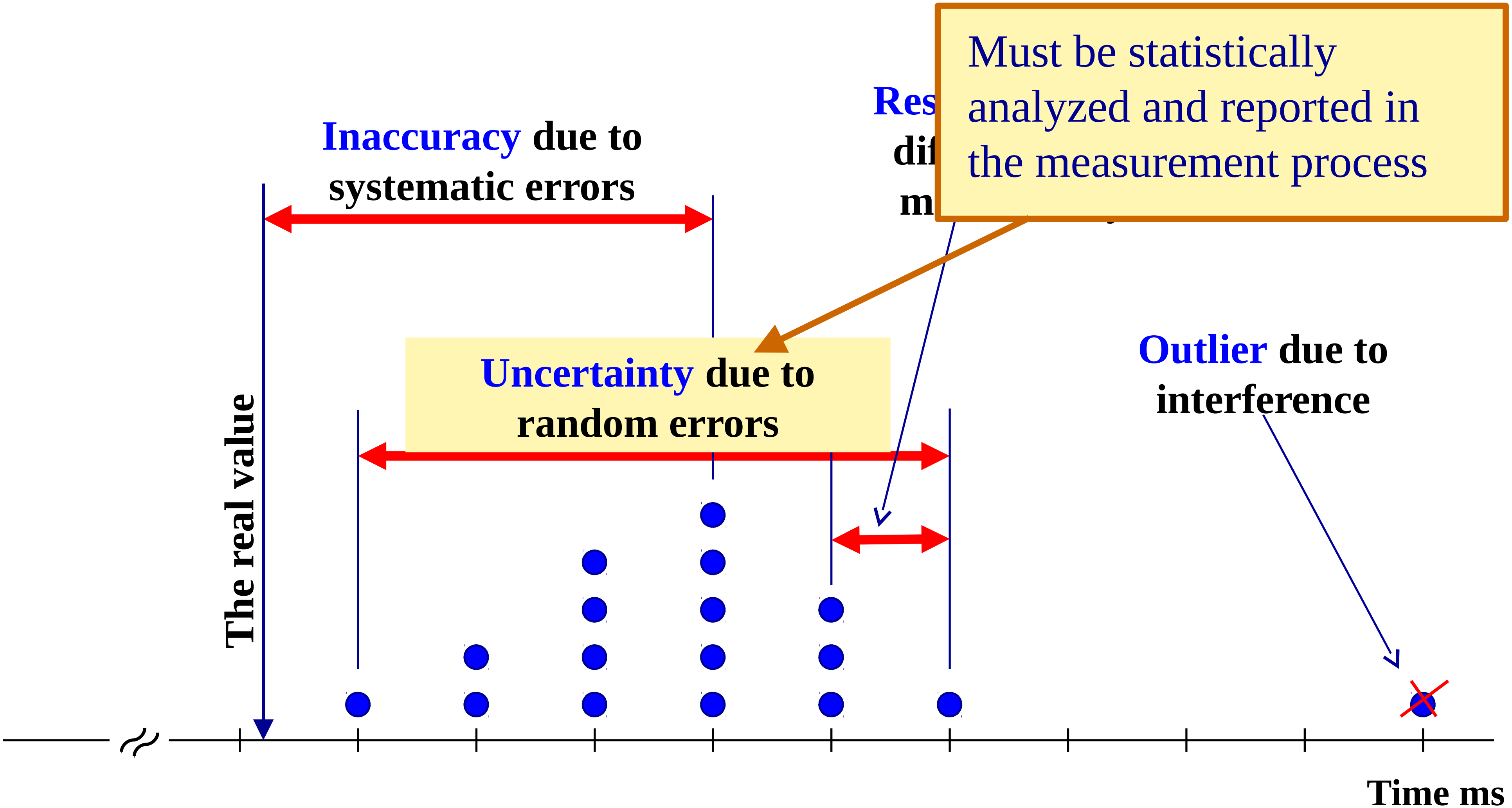
- **Changes in the conditions of the measurements (experiment environment, handling techniques, etc.).**
  - For examples, small changes in the load of a computer, cache state, available network bandwidth, etc., in the different measurements.
  - Quite often, the small changes in the experiment environment are analyzed statically as random uncertainties.
  - Extreme cases lead to outliers (should be ignored or investigated)

Summary



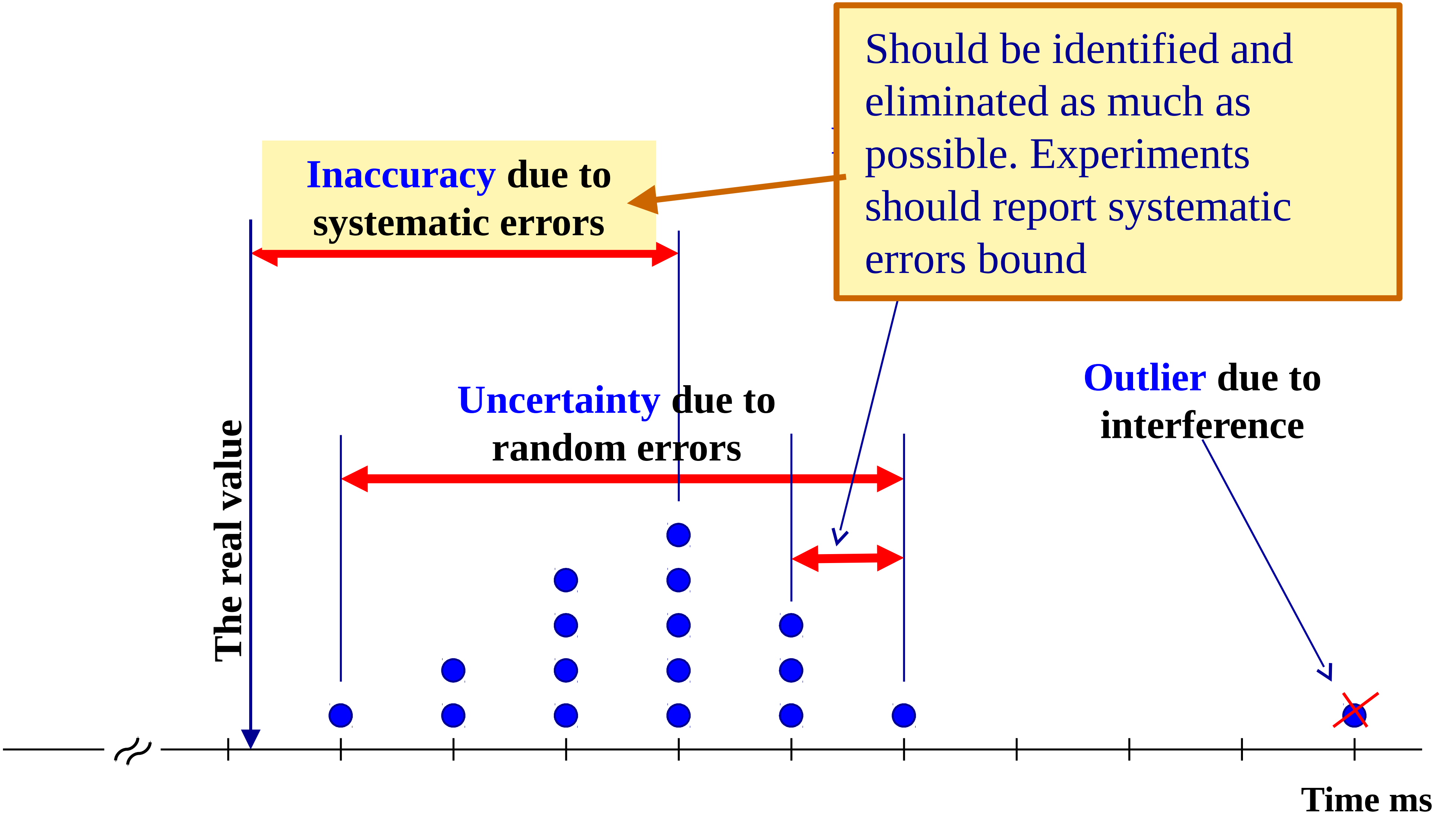
(adapted from Dror Feitelson, HU, slides)

Summary: what should we do?



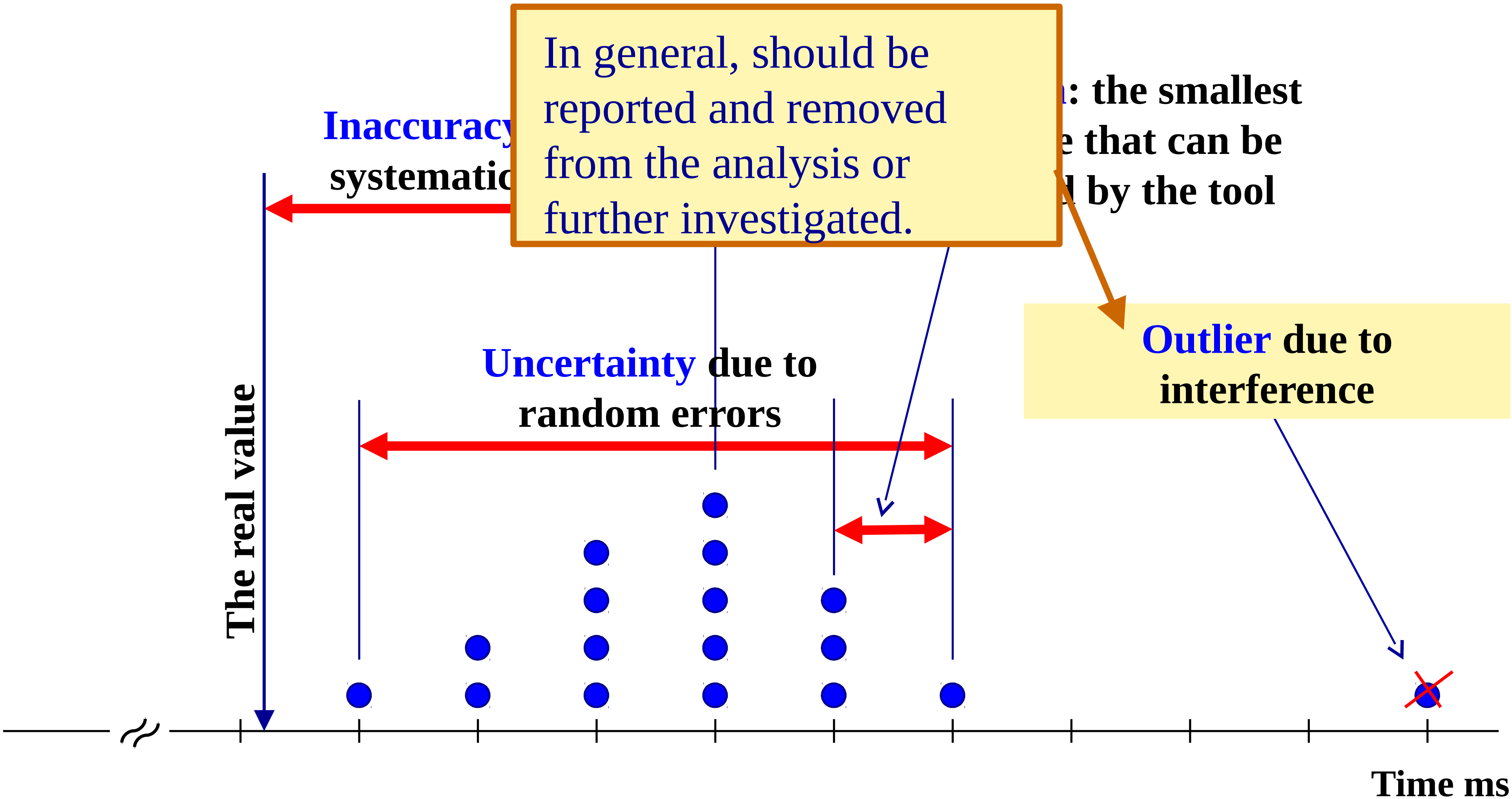
(adapted from Dror Feitelson, HU, slides)

Summary: what should we do?



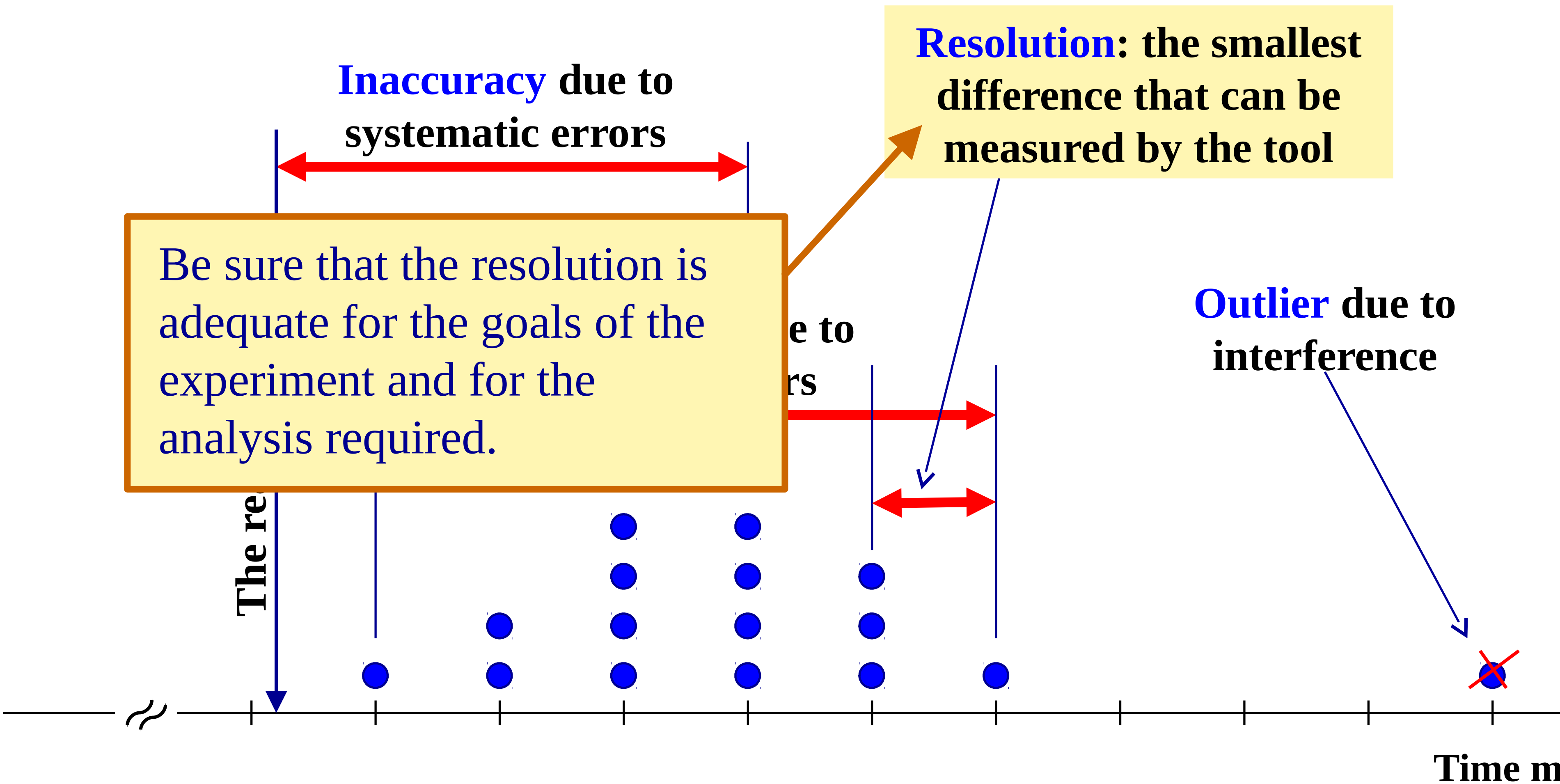
(adapted from Dror Feitelson, HU, slides)

Summary: what should we do?



(adapted from Dror Feitelson, HU, slides)

Summary: what should we do?



(adapted from Dror Feitelson, HU, slides)

Expected value and sample mean

- The measured values ($x_1,\ldots,x_n$) can be seen as a random sample from a population: measured values are values of a random variable X with an unknown distribution.

- The most common index of central tendency of X is its mean E[X], also called expected value of X.

$$E[X] = \sum_x x \Pr(X = x)$$

- The sample mean (arithmetic mean) is an estimate of E[X]

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(See David Lilja, 2000)

Indices of central tendency

- Sample mean
  - Use when the sum of all values is meaningful
  - Outliers can have a large influence on the computer mean value
- Median
  - The value such that ½ of the values are above, ½ are below.
  - If n is even, median is the mean of the two middle values
  - Less influenced by outliers
- Mode
  - The value that occurs most often (may not be unique)
  - Use when values represent categories.

Other types of means

- Harmonic mean
  – To summarize rates (e.g speedup)

$$\bar{x}_{\mathrm{H}} = \frac{n}{\sum_{i=1}^{n} 1/x_i}$$

- Weighted means

The previous definitions of means assume all measurements are equally important. If not the case, one can use weights to represent the relative importance.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} w_i \, x_i$$

$$\bar{x}_{\mathrm{H}} = \frac{n}{\sum_{i=1}^{n} w_i/x_i} \qquad \sum_{i=1}^{n} w_i = 1$$

Indices of dispersion

- Sample variance

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n\text{-}1}$$

- Standard deviation
  - Square root of variance
  - Same units as mean

A model of errors

- Gaussian model of errors
  - Assume that a single source of random error can change the value measured for x by $+E$ or $-E$ with equal prob.

Error	Measured value	Probability
$-E$	$x - E$	½
$+E$	$x + E$	½

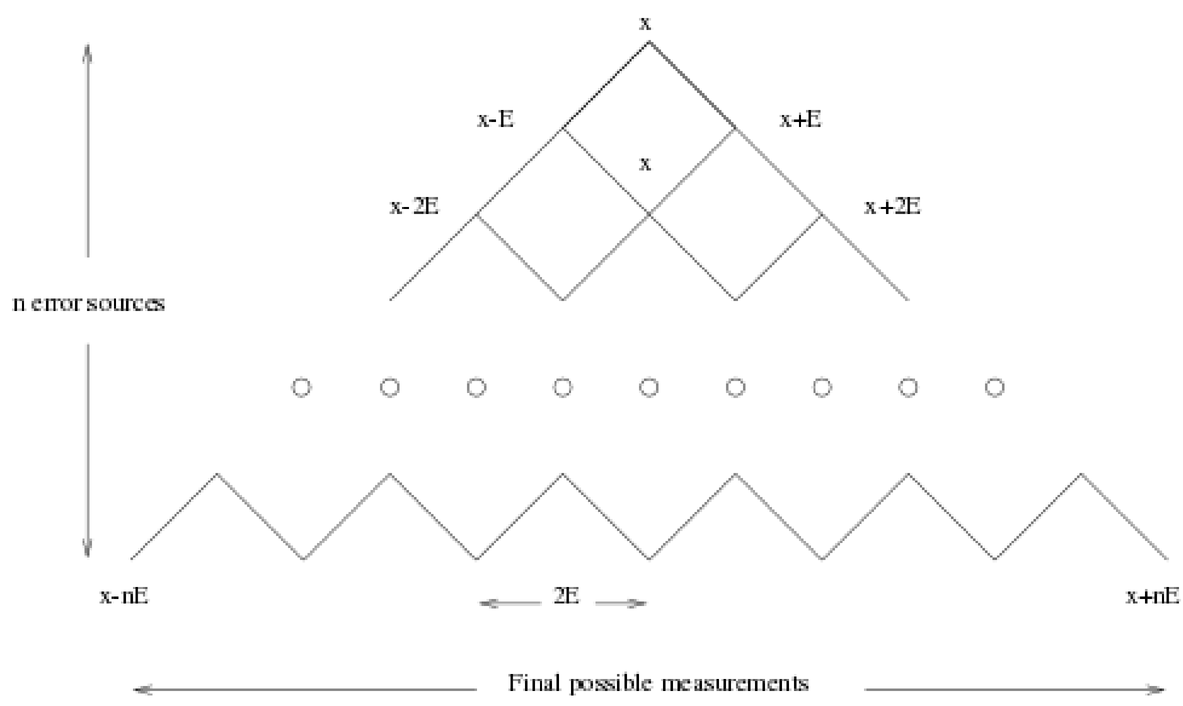
A model of errors

- Gaussian model of errors
  - Assume that **two** sources of random errors can change the value measured for x by $+E$ or $-E$ with equal prob.

Error 1	Error 2	Measured value	Probability
$-E$	$-E$	$x-2E$	¼
$-E$	$+E$	x	¼
$+E$	$-E$	x	¼
$+E$	$+E$	$x+2E$	¼

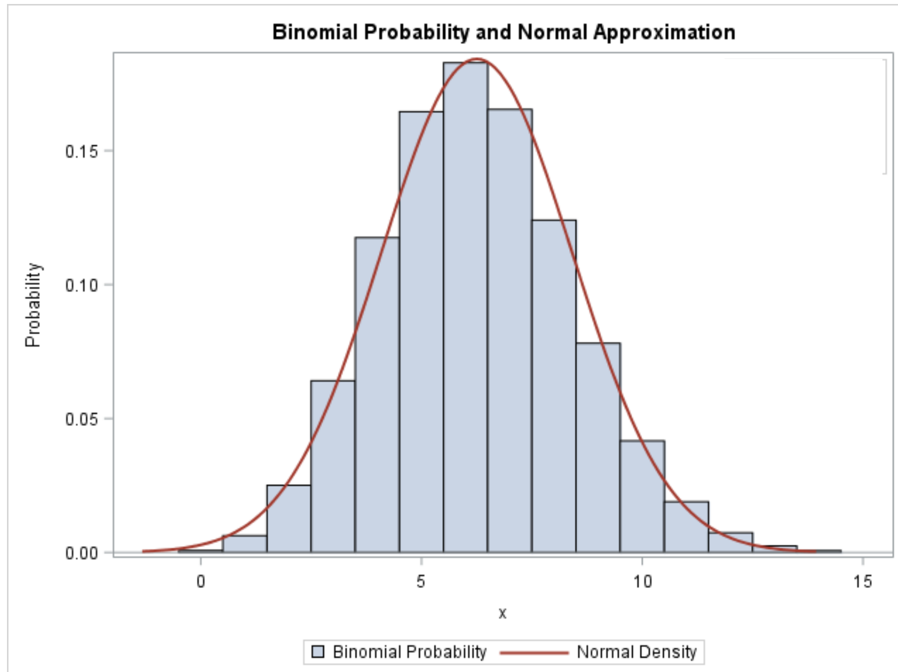
A model of errors

- Gaussian model of errors
  - Assume that **n** sources of random errors can change the value measured for *x* by +*E* or –*E* with equal prob.



A model of errors

- Gaussian model of errors
  - The probability of obtaining a certain measurement is proportional to the number of paths that lead to that measurement.
  - This produces a binomial distribution for the possible measurements – for large *n* this distribution approximates a Gaussian distribution.



This assumption is the basis of parametric statistics.