The DIN/ISO definition and a measurement procedure of software effiency

Dr. W. Dirlewanger (Prof. i. R.)
Dept. Mathematik/Informatik
Kassel University
34246 Vellmar Deutschland
Email: performance-we@t-online.de

Ladies and Gentleman !

My topic is the method for

Performance measurement
and
SW efficiency measurement

as described in the international ISO standard 14756
and
its predessessor the national German standard 66273

Surely you have heard from it.

>>>>>>>>>>>>> But if you are not familiar with it:
Here is a short introduction

>>>>>>>>>>>>>< May be yo u know it but don' use it:
Here I will show advantages

>>>>>>>>>>>>> In case of already using it:
I hope I can give you some additional ideas

Contents

What is SW-“Perf“ ?

Comp-Perf-Measurement

1. Standards & Research

- National and international Standards:
  For instance screws, measures,...

- Work of a standardisation commitee (typical):
  x) Look for existing solutions
  x) Decide which is the most fitting one
  x) Propose it for Standard

- Interest of „research and development“ in standards ???

>>>>> Normally very few <<<<<

What‘s about DIN66273/ ISO 14756 ?

-DIN started a (national) standardization project:
Measurement and rating of computer performance

-Many computer manufacturers and users
were interestet

-Working group: among others:

Mainframe: IBM, Siemens, Comparex,
Unisys, ...

Mid size: HP, Nixdorf, DEC, ...

Universities: Kassel, Neubiberg, Tübingen

Users : German Telekom, ...

>>>>>>>> (up to 30 Persons) <<<<<<

- Starting: Existing methods/measures ?

for instance

- Runtime of a Job
- Mean value of runtimes of a job set
- Reciprocal of runtime
- Number of instructions per time unit (MIPS)
- Loops per time unit
- Quotient of total runtime of a set of benchmarks (actual system and reference system)
- customer Umfrage-Verfahren: individual estimation (not measuremernt) of MIPS
- OLTP: no methods at all

>>>>>>> poor poor >>>>>>>>>>>

Most fitting Performance measure ? none

Most fitting Measurement method ? None

The working group shiftet

Commetee >>>> Resarch group

- Decision: No attempt to declare an existing method
  for a standard
  but develope a complete new method

- Goals: x) enduser oriented,
  x) fitting for
    – all IP systems
    – all computer architectures/structures,
    – systems of any size

Result: 2 revolutionary standards

DIN series 66273 (1991 ff.)
Performance Measurement

ISO/IEC 14756 (1999)
ISO took over the 66273

- replace the oldfashioned definitions and measurement methods.

- The new measurement method is a new basis for what colloquially is called Software Performance

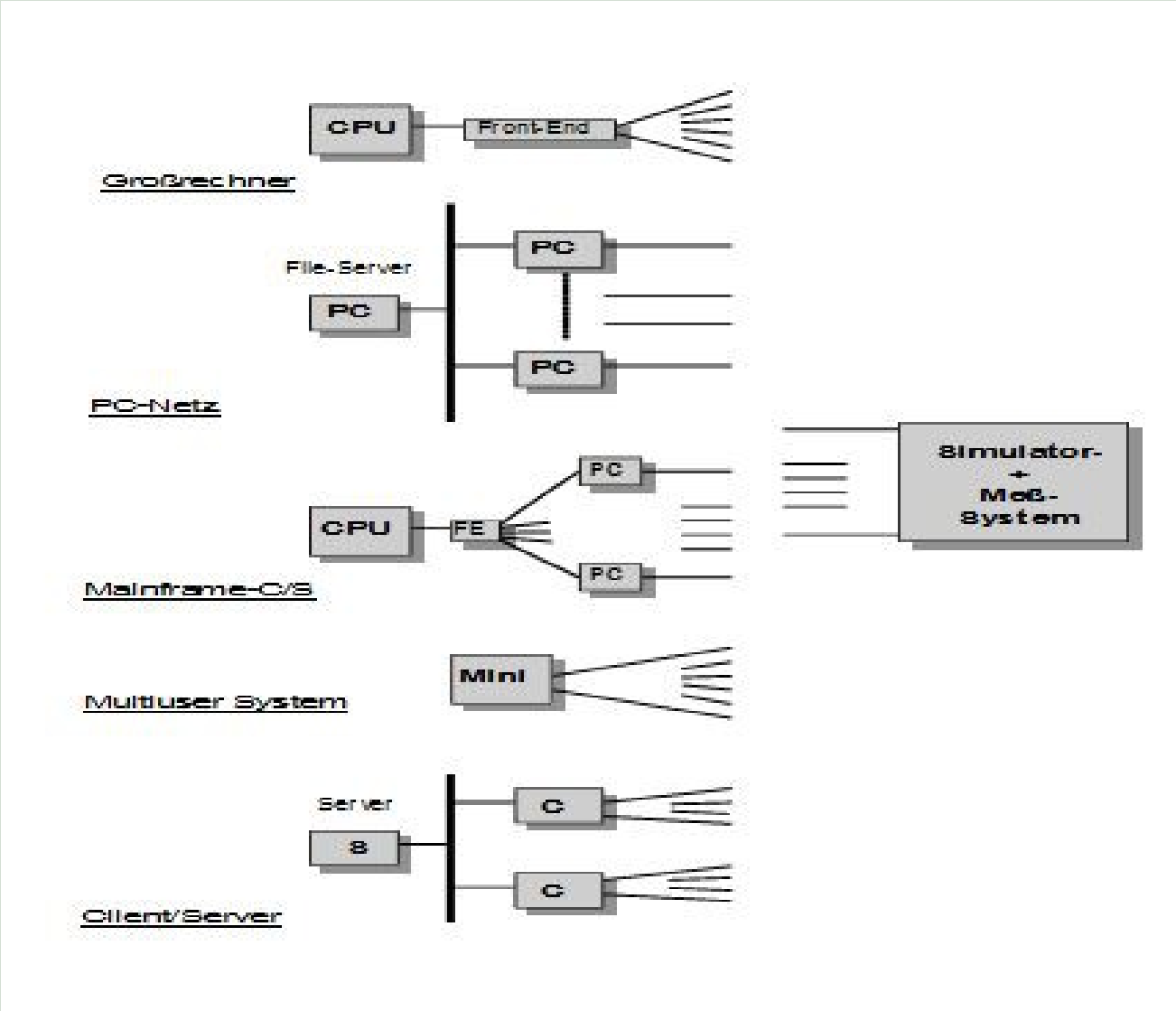
- ISO added this topic and called it „SW efficiency“

2. Special qualities of the ISO 14756 method

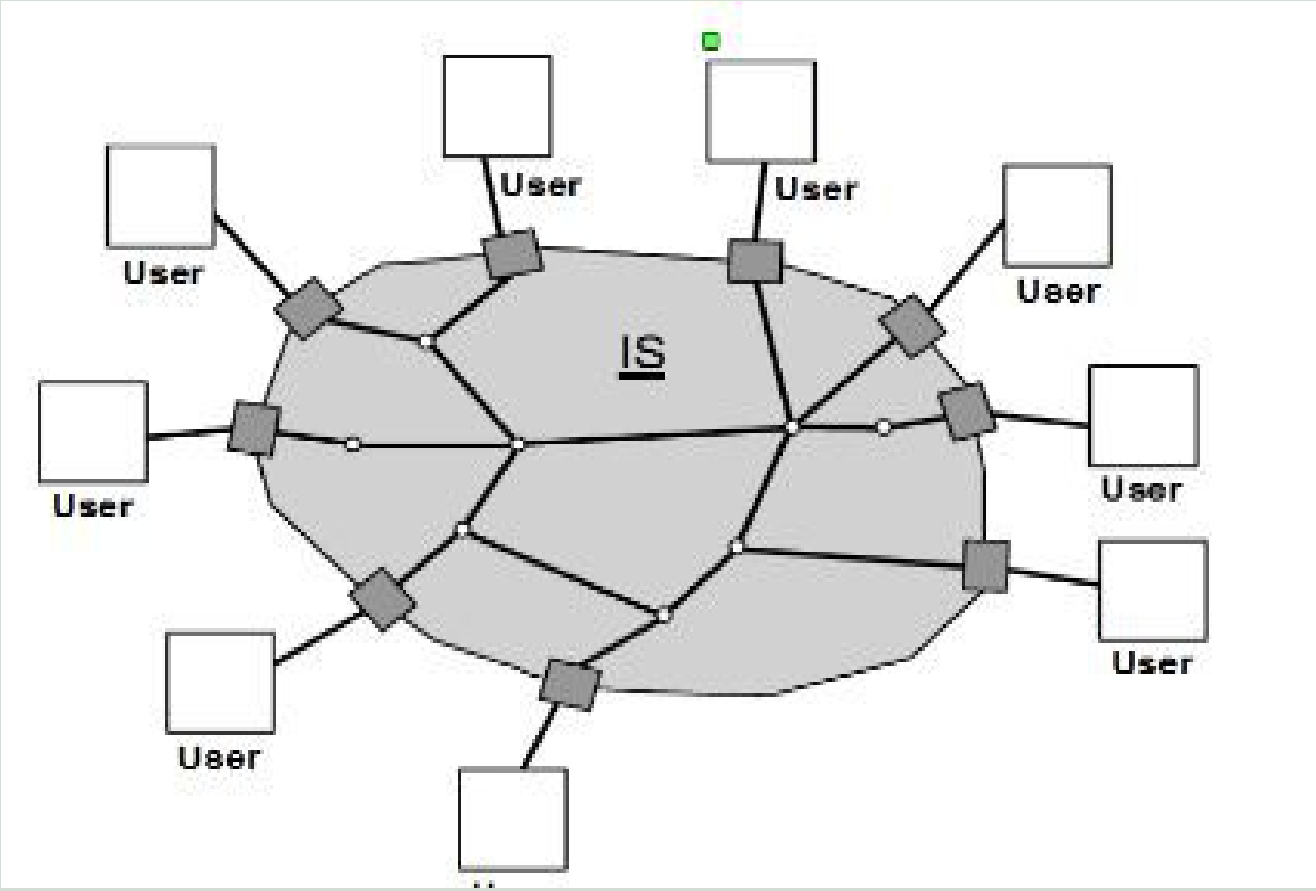
- Arbitrary system for a SUT
- Independence of RTE's manufacturer
- Control of correct work
- Nearly every benchnark can be represented in ISO form
- Also component tests can be rewritten in ISO form
- Emulated users can be human beeings or machines
- Forgery proof by random task stream
  (microscopic: random, macroscopic: deterministic)
- Reproducibility of measurement results
- Applicable also to simulated SW
- High precisision of measurement results

3. The ISO measurement method

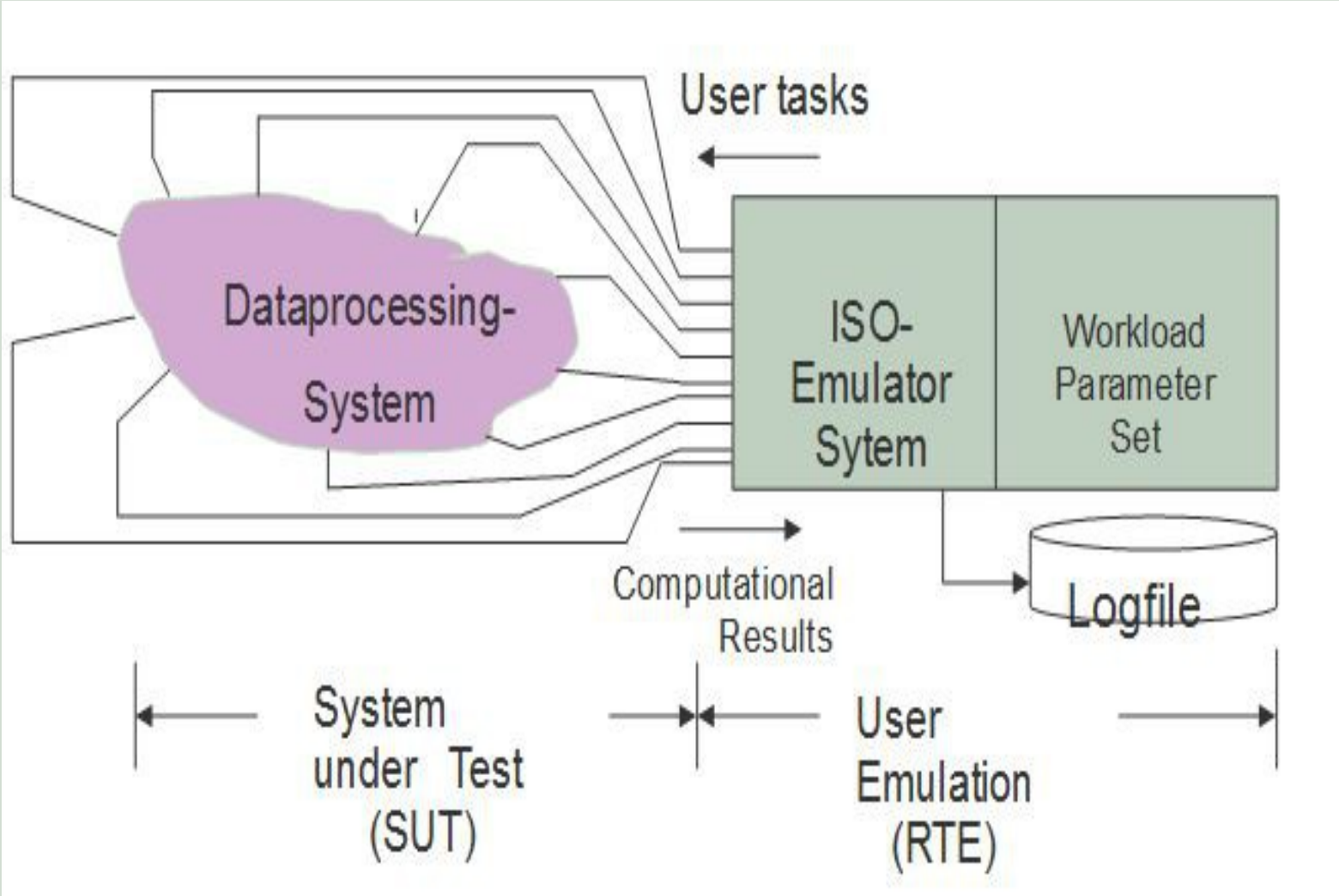
Any type of DP System：



A SUT in real operation



ISO measurement: RTE replaces the users



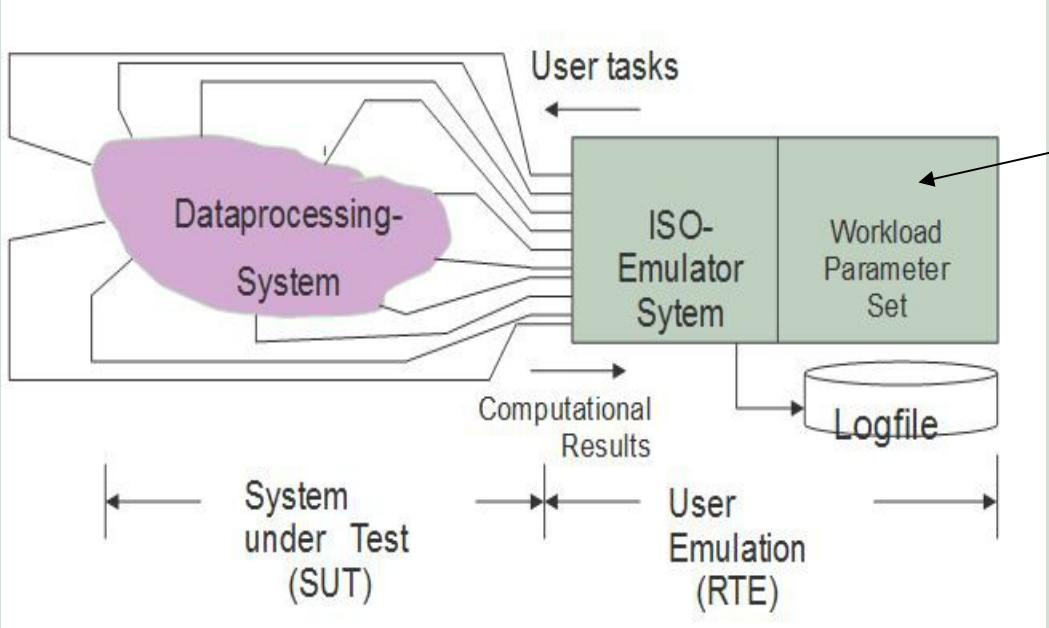
The ISO workload:

1. Application programs
2. OS command procedures
3. User data stored in the SUT
4. All computational results
5. Parameters for controlling
   a) correct work of the RTE
   b) correct work of the SUT
   c) statistical significance of measurement results

- and -

6. - last not least – the WPS

Measurement configuration



WPS

Workload Parameter Set

RTE replaces the real users.

- RTE table driven by the WPS (workload parameter set)

1. Basic parameters:

Number n of user types
Number of users of each type $N_{user}(1),..., N_{user}(n)$
Number w of activity types
Number p of timeliness functions
Number m of task types
Number u of task chain types

2. Definitions of the w activity types (i.e. the elementary end user actions)

each described by :
Input string or mouse pointer action, rules for input variation if there is so.

3. Definitions of the m task types, each defined by a triple:

(Activity type + WAITmode + TF)

- wait mode WAIT/NOWAIT for the result of the actual task and

- Timeliness function:
  Enduser's requirements for completing the task

- Example

At least 80% within 2 sec
Maximal 15% within 6 sec
maximal 5% within 15 sec
none longer than 15 sec
A upper limit is mandatory

4. Task chains

a) Definitions of the u task chain types:

Chain length (number of tasks) and task type sequence

b) Definition of each of the user types by the n x u matrix of relative chain frequencies q(i,l) where i is the current number of the user type and l is the current number of the Chain type.

5. Statistic parameters of the (random to be created) think times

of the users

Firstly: matrix of n x m think time mean values

(Remark: Think time is task preparation time.)

Secondly: matrix of n x m think time standard deviation values

6. Criteria of statistical significance of the measurement result

a) ALPHA (confidence coefficient)

b) D_{rel} (half width confidence interval)

Surprising:

Assume a SUT which executes all tasks so fast that all timeliness functions are just fulfilled and none faster:

Throughput $B(j)$ and Mean responsetimes $T_M(j)$

$j=1, 2, \ldots, m$

computed directly from WPS without any measurement:

$B_{Ref}(j)$

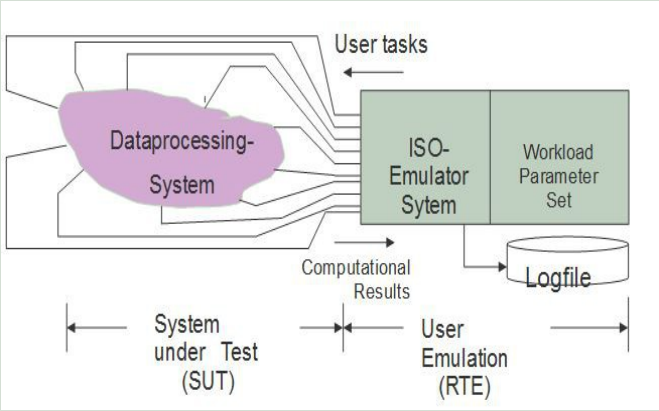
throughput reference value(s)

$T_{Ref}(j)$

response time reference value(s)

- This is the so-called theoretical reference machine

Measurement



<u>Steps:</u>

- Install applications in the SUT
- Load workload parameter set (WPS) into RTE
- Run and record logfile. 3 Phases:
  Stabilisation phase -- Rating interval -- Supplementary run
- Store computational results
- Checking correctness
  (RTE: correct work and statistical significance of random variables;
  SUT: correct and complete computational results)
- Testing statistical significance of the results
- Analysis of recorded data and computation of performance
  values and rating values

4. Results of a measurement

4.1) Measured performance values

P is a triple of vectors $\boxed{P = (B, T_{ME}, E)}$ 3 x m values

4.2) Ratings

Compare measured values to those of the "theoretical reference machine":

R is a triple of vectors: $\boxed{R = (R_{TH}, R_{ME}, R_{TI})}$ 3 x m values

m is the number of task types

Only if all of the 3 x m rating values are not less 1 the SUT satisfies the timeliness requirements of the user entity.

.

4.1.(cont.) Formulea

Performance P computed from the recorded logfile

P triple of vectors: $\boxed{P = (B, T_{ME}, E)}$

(total) throughput vector $B = (B(1), \ldots, B(m))$.

B(j) is the mean number of tasks of the j-th task type sent from the RTE to the SUT per time unit.

execution time vector $T_{ME} = (T_{ME}(1), \ldots, T_{ME}(m))$.

$T_{ME}(j)$ is the mean execution time of tasks of the j-th task type.

timely throughput vector. $E = (E(1), \ldots, E(m))$

E(j) is the mean number of tasks of the j-th task type which were timely executed by the SUT per time unit.

4.2 (cont.) Rating of the measured performance

Compare measured values to those
of the "theoretical reference machine":

$B(j)$ to $B_{Ref}(j)$ throughput mean values

$T_M(j)$ to $T_{Ref}(j)$ mean response times

$E(j)$ to $B(j)$ timeliness

$j = 1, 2, \ldots, m$

<u>Formulea</u> (rating values)

Throughput rating vector $R_{TH} = (R_{TH}(1),..., R_{TH}(m))$

with

$$R_{TH}(j) = B(j) / B_{Ref}(j)$$

$B_{Ref}(j)$ is the throughput of the j-th task type of the so called theoretical reference machine.

Execution time rating vector $R_{ME} = (R_{ME}(1),..., R_{ME}(m))$

with

$$R_{ME}(j) = T_{Ref}(j) / T_{ME}(j)$$

$T_{Ref}(j)$ is the mean execution time of tasks of the j-th task type. of the so called theoretical reference machine.

Timely throughput rating $R_{TI} = (R_{TI}(1),..., R_{TI}(m))$ is the timely throughput rating vector

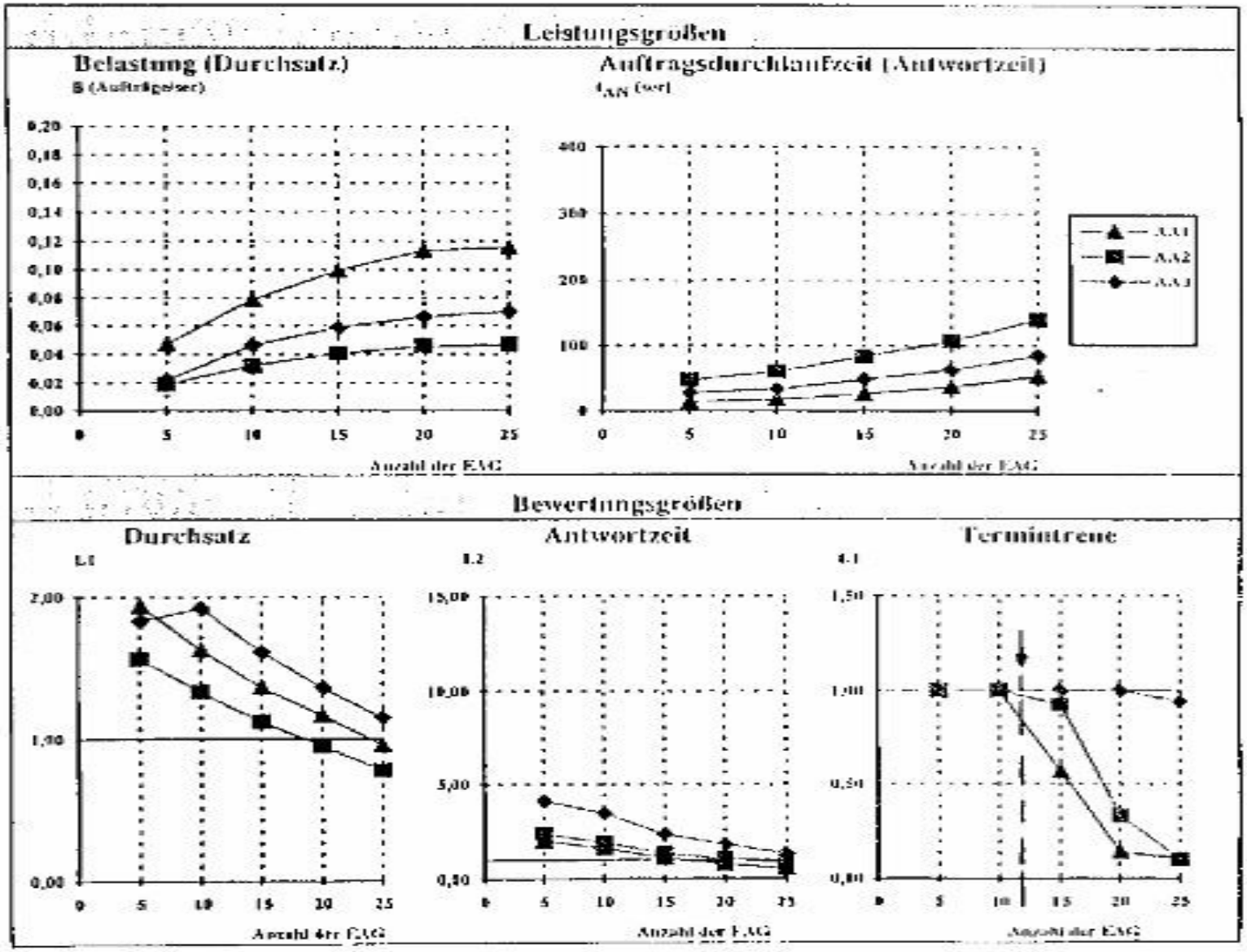
with

$$R_{TI}(j) = E(j) / B(j)$$

Example 0.A ISO measurement and rating of a mainframe (measurement series, 5 to 25 users)

Per-
For-
mance

Rating



Only if all of the 3 x m rating values are not less 1
the SUT satisfies the timeliness requirements of the user entity.

Elsewhere the system has to be rejected due to insufficient response times.

5. SW performance ? (Finding a term)

SW qualities: storage usage,
Changeabiltay,
maintainability,...
- and -
runtime qualities

SW has not a property „speed“ or „performance“

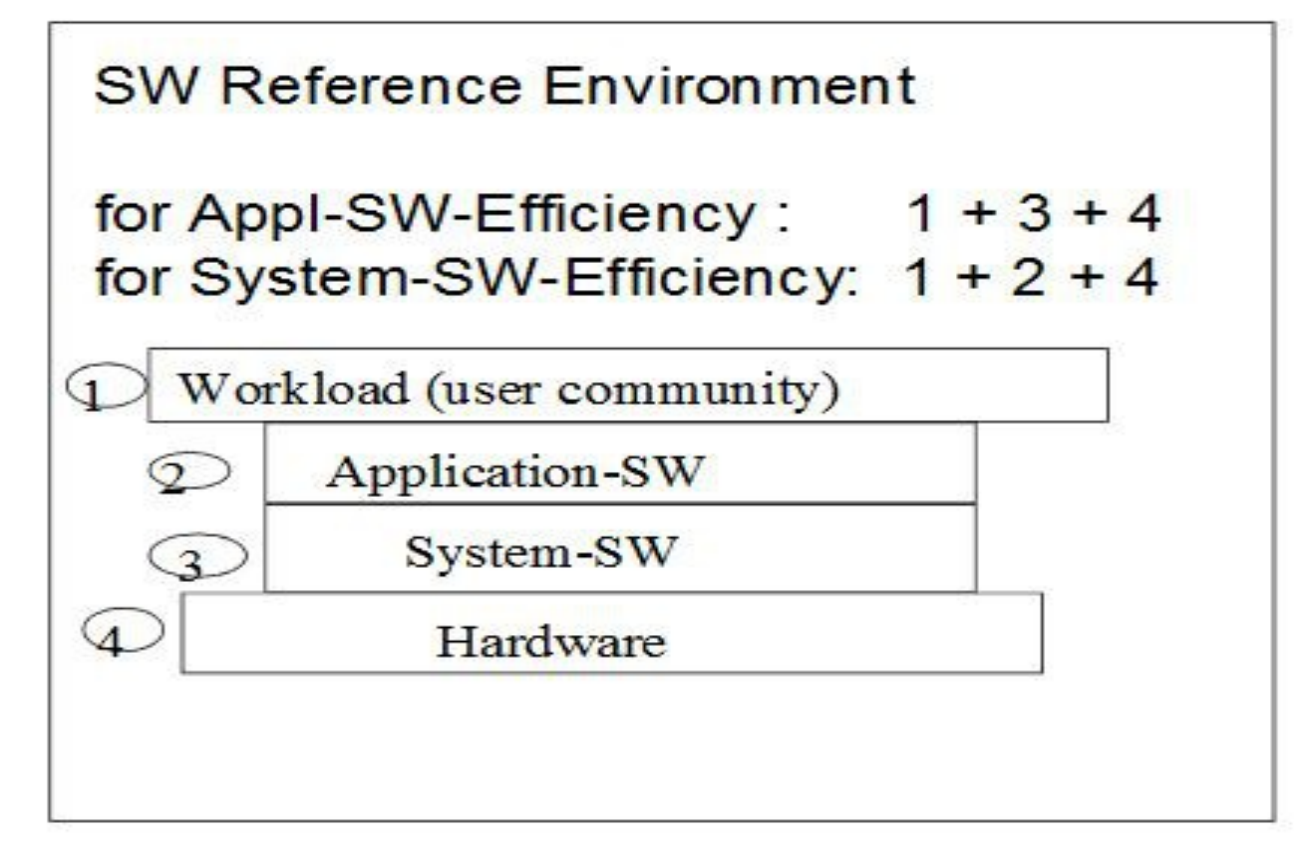
SW consists of sequences of (machine- or HLL-) instructions
to be performed by a CPU.

Fast CPU >> short time for a user task | Slow CPU>> Long time

Finding a term describing „performance of SW:

- Take two Implementations.
- Install in the same Reference Environment
- Measure performance P1, P2

Compare P1 to P2 >>> SW efficiency



4.2.1):
Throughput efficiency values

$I_{TH}(x) = B_2(x) / B_1(x)$ x=current number of task type

Values in example 0.A (below):

x=1

(for n=10 users) (for n=15 users)

0.130 / 0.080 = 1.63 0.125 / 0.100= 1.25

x= 2

(for n=10 users) (for n=15 users)

0.058 / 0.035 = 1,66 0.080 / 0.410 = 1.95

x = 3

(for n=10 users) (for n=15 users)

0.058 / 0.450 = 1,29 0.081 / 0.060 = 1,35

4.2.2)
Mean execution time efficiency values
Values in on example 0.A (below):

$$I_{ME}(x) = T_{ME1}(x) / T_{ME2}(x)$$

x= current number of task type

x=1

(for n= 10 users)	(for n=15 users)
14.5 / 2.10 = 6.90	20 / 4.5 = 4.44

x=2

(for n= 10 users)	(for n=15 users)
61.0 / 21.9 = 2.78	80 / 20.5 =3.90

x=3

(for n= 10 users)	(for n=15 users)
39.0 / 13.5 = 2.89	50 / 26 = 1.87

4.2.3)
Timely throughput efficiency values

Values in on example 0.A (below):

$$I_{TI}(x) = E2(1) / E1(x)$$ x = current number of task type

x =1

(for n= 10 users)	(for n=15 users)
0.130 / 0.080 = 1.51	0.125 / 0.060 = 2.08

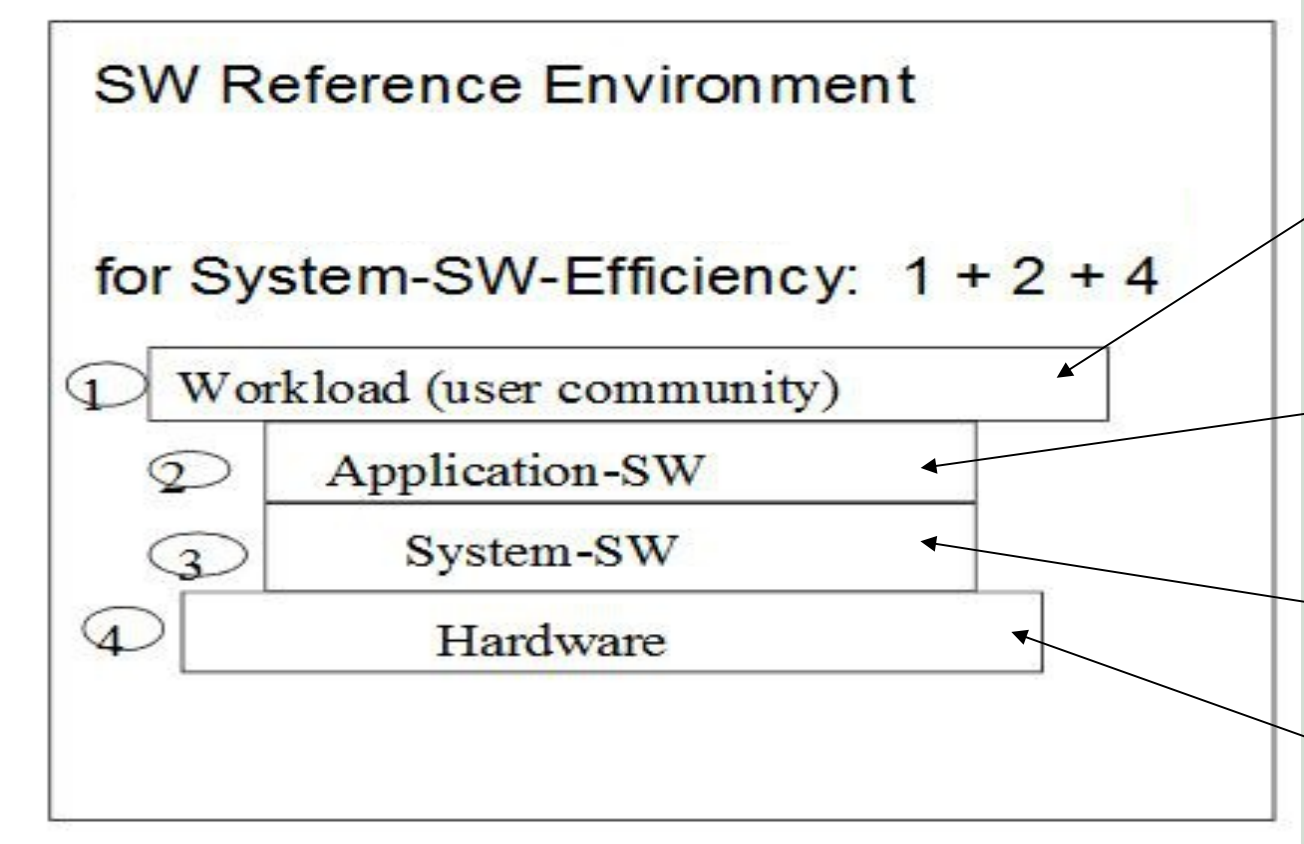
x =2

(for n= 10 users)	(for n=15 users)
0.058 / 0.035 = 1.66	0.080 / 0.038 = 2.34

x=3

(for n= 10 users)	(for n=15 users)
0.058 / 0.045 = 1.23	0.080 / 0.060 = 1.43

6. SW efficiency Example 1: System-SW efficiency in computer center operation



Up to 30 users

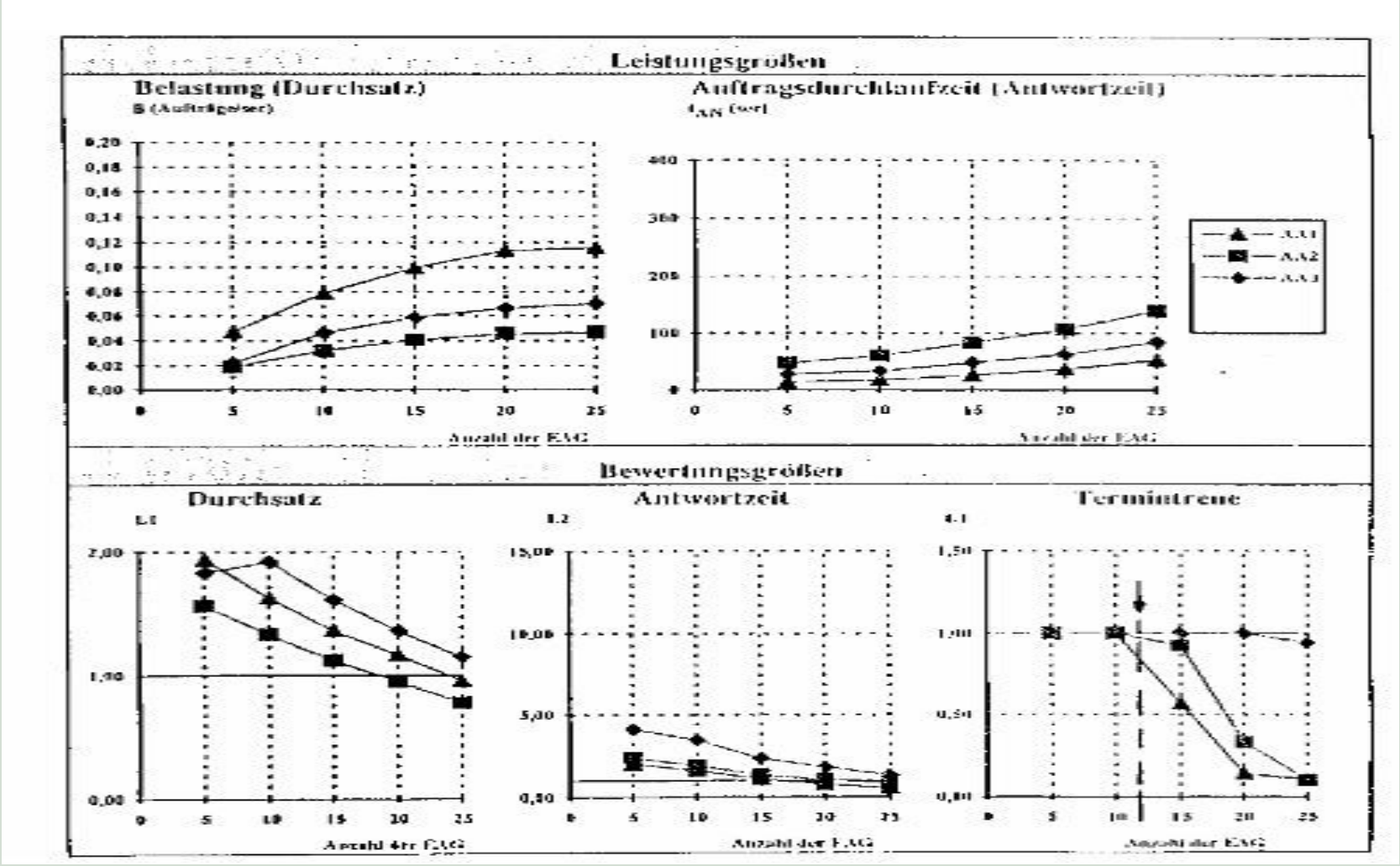
Comp center operation (on-line + batch)

SW efficiency OS2/OS1 = ?

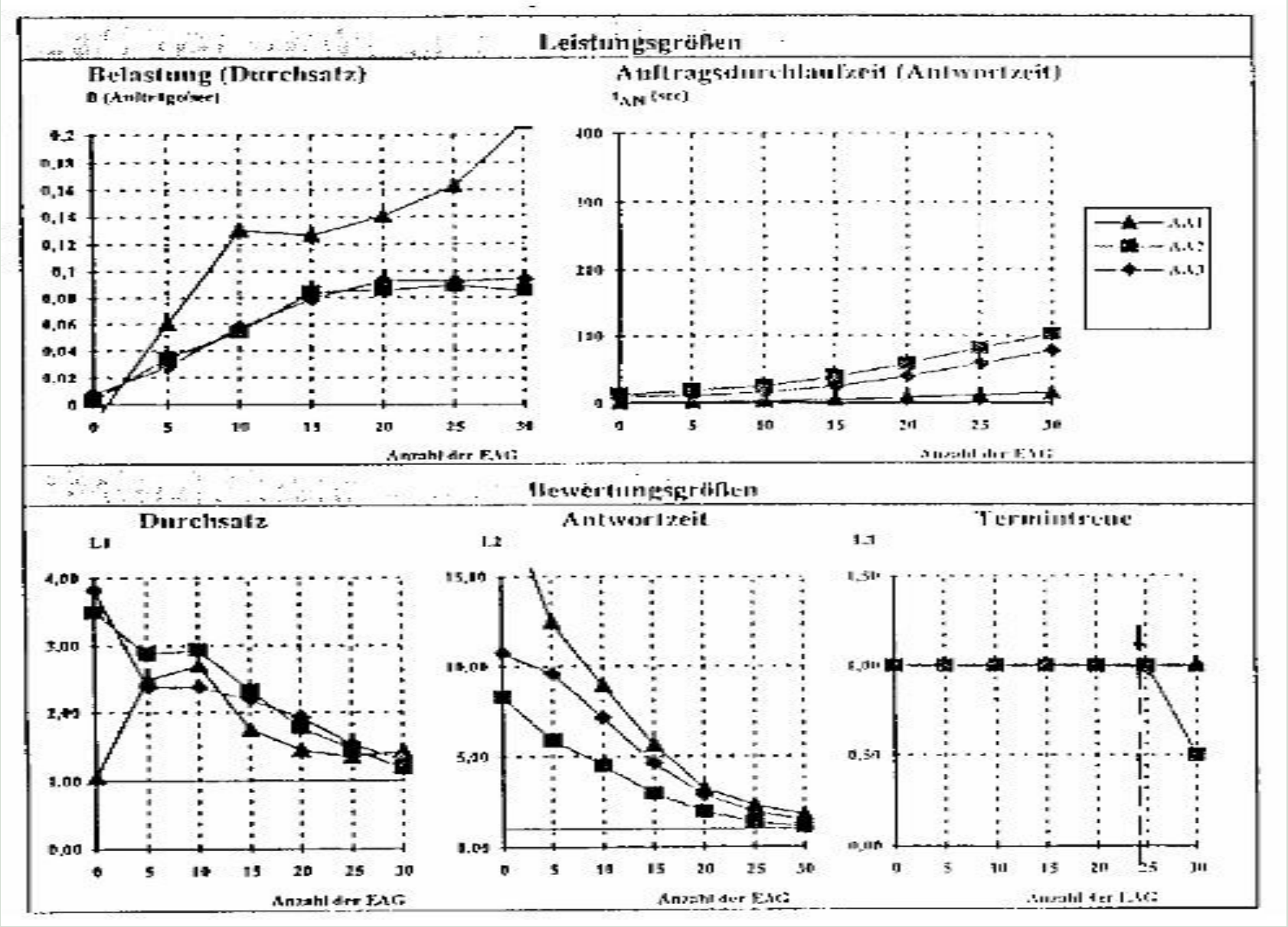
Mainframe

6. SW efficiency Example 1

a) Mainframe using OS1



b) The same mainframe but using OS 2



Summary:

a) n =10 users

Throughput oriented efficiency values: around 1.5

Response time oriented efficiency values: nearly at 3.0

b) n =15 users

Throughput oriented efficiency values: around 2.0

Response time oriented efficiency values: nearly at 4.0

Remarks:

a)Those values show that n=10 and n=15 are different SW environments. Having a greater number of users the difference in SW efficiency of OS2 to OS1 increases.

b)But be aware that the difference in case of 15 users is meaningless. The reason: OS1 does not serve 15 users timely. The maximum number of timely served users is 12.

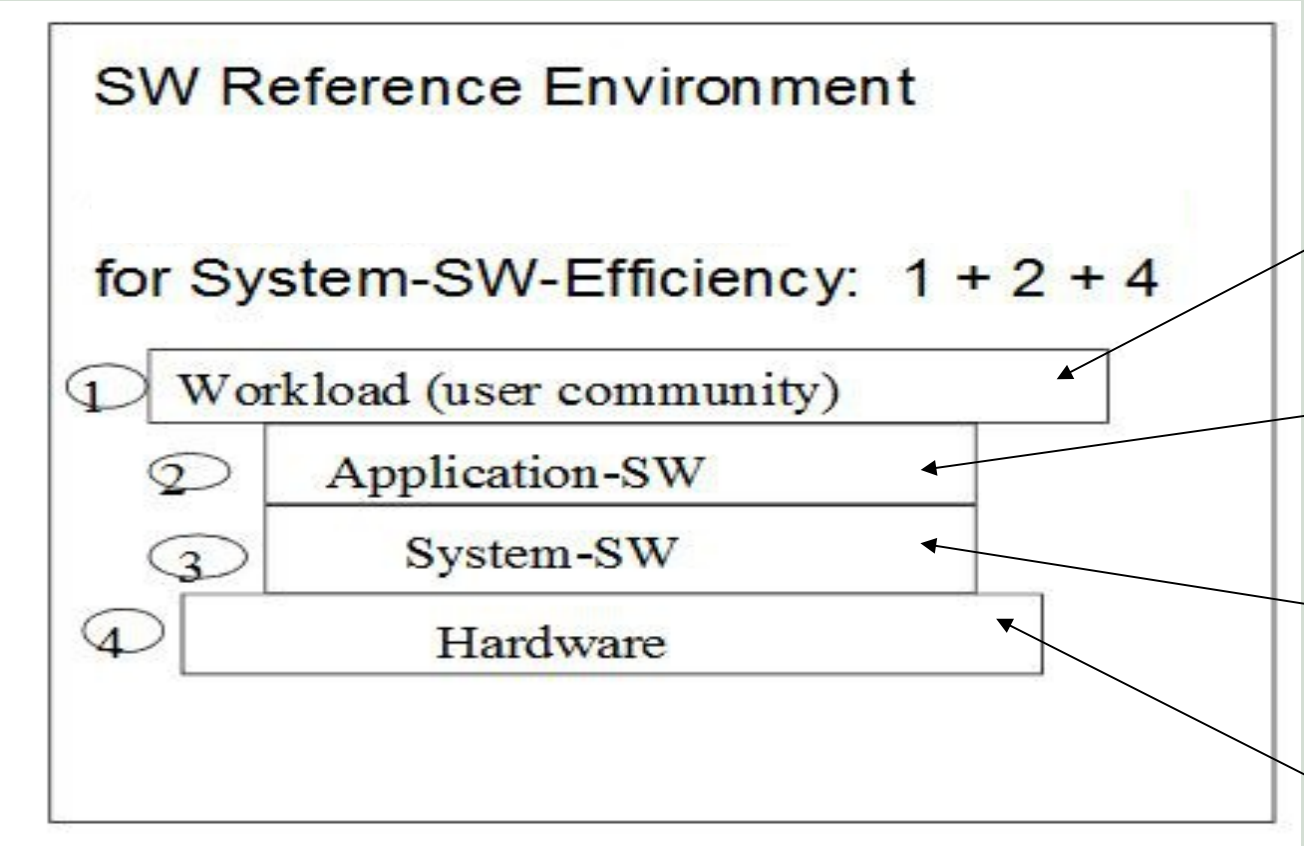
c) Additional software efficiency measure:

$$I_{max} = N_{max}\ 2 / N_{max}\ 1 \qquad 24 / 12 = 2 \quad .$$

Final result :Using OS 2 instead of OS1 the computer system serves (about) 100% more users timely

I. e. 100% more SW efficiency in this environment

7. SW efficiency Example 2: System-SW efficiency in OLTP operation



Up to 200 users

SAP/R2

SW efficiency
OS2/OS1 = ?

Main-
frame

Example 2

The hardware and operating systems are the same as in example 1.
But the application software is SAP/R2 instead of the software contained in the (modified) ISO workload CC1 of example 1.

Used were 4 parts of the SAP software:

RF (Finances) : 40% of the users
RM-MAT (Materials management): 30% of the users
RM-PPS (Production) : 10% of the users
RV (Sales) : 20% of the users

An ISO type workload was developed. The workload parameter set (shortened):

4 user types
110 activity types
110 task types
21 chain types
3 timeliness functions (mean values 3, 6 and 36 seconds)

Measurement results:

a) n =5,10, 15, 20 ,25, 30 users

Throughput oriented efficiency values: …….

Response time oriented efficiency values: .....

b) Additional software efficiency measure:

Measured N_{max} values:

Operating system 1: $N_{max1} = 110$

Operating system 2: $N_{max2} = 170$

$I_{max} = N_{max2} / N_{max1}$ 170 / 110 = 1.55

Final result :Using OS 2 instead of OS1 the computer system serves (about) 55% more users timely

I. e.: OS2 has 55% more SW efficiency in <u>this</u> environment

8. ISO 14756 and simulated SW

- Applicable also to simulated SW

- Recommendation: Use ISO 14756 from first steps to end of project (from birth to death)

- ISO workload specification is a template for the SW environment

  >>> high SW efficieny
  >>> saving man power
  >>> usefull for all phases of a SW project and for all members

9. Final remarks

- ISO 14756: Universal method
  High precision
  Results reproducible

- Method complex.Time is needed for understanding
  Example: Understanding backfeed „thinktimes – throughput“;
  users often „lacy“;

- ISO-RTE were implemented (by Comp-Cs, Manufactureres,....)
  (Demo in Ref [4] Professional Implementation in Ref [5] Proceedings)

- The ISO-workload description is a template for
  „Environment when measuring SW efficiency“

This was a look into the method for

Performance measurement
and
SW efficiency measurement

as described in the international ISO standard 14756 and its predessessor the national German standard 66273

>>> If you were not familiar with it:
Here was a short introduction

>>> If you know it but don't use it:
Here I showed advantages

>>> In case of already using it:
I hope I could give you some
additional ideas

Thank you
for listening