CLOCK SYNCHRONIZATION

H KOPETZ

Rapporteur: M J Elphick

TU Wien		rt.3 2.3

Different times to consider in a DRS:

(1) local "Political Time"

(2) Universal Time Coodinated (UTC)

(3) External Physical Time (TAI)

(4) Internal Physical Time

(5) Approximate global time

(6) Local real time clock

Properties of a time base in a distributed real time system:

- metric of physical second

- chronoscopic, i.e. can be used for the measurement of small intervals at any point in time

- bounded accuracy of synchronization

- fault tolerant

Internal synchronization:

Synchronization of the times of the local real time clocks in order to generate the (approximate) global time.

Synchronization Accuracy Δ_{int}

Granularity: n_g

External synchronization: Δ_{ext}

Synchronization of the approximate global time with the external time standard.

"Reasonable" Timebase:

$$|\Delta_{int}| < n_g < |2\Delta_{int}|$$

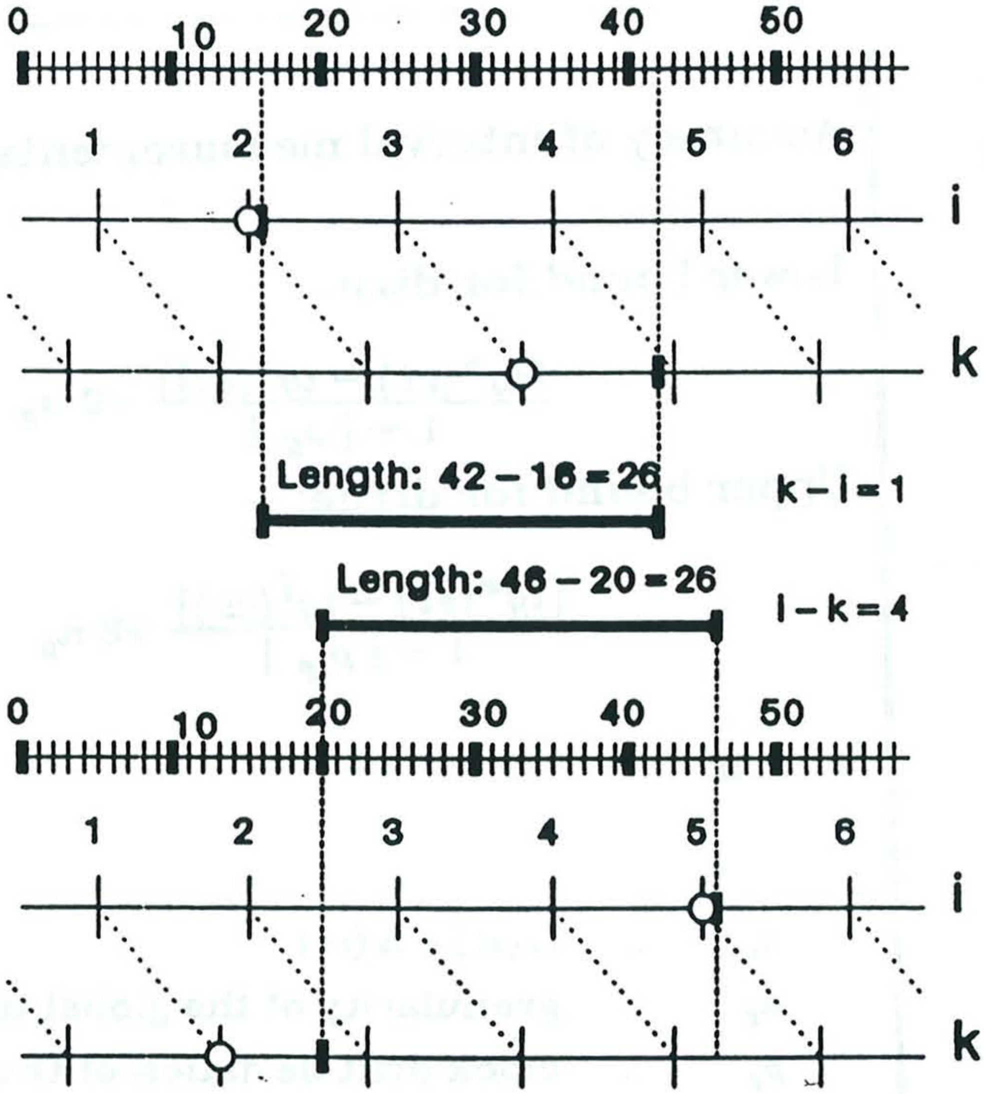
Given a 'reasonable' global digital time base and two events, e1 and e2, where

$$tg(e1) - tg(e2) = n$$

then:

$n \leq -2$	e1 definitely occurred before e2
$\lvert n \rvert < 2$	e1 and e2 occurred about at the same time, we do not know which one was first
$n \geq +2$	e1 definitely occurred after e2

TU Wien		rt 6b R.7



Accuracy of interval measurements <es,et>

Lower bound for dtrue:

$$\frac{[tg^k(et) - tg^i(es)]}{1 + |\rho_g|} - 2\, n_g$$

Upper bound for dtrue:

$$\frac{[tg^k(et) - tg^i(es)]}{1 - |\rho_g|} + 2\, n_g$$

where:

d_{true}	..	$ts(et) - ts(es)$
n_g	..	granularity of the global time
ρ_g	..	clock drift deviation of the ensemble
es	..	start event
et	..	termination event

TU Wien | rt.7 | R.9

Internal clock synchronization:

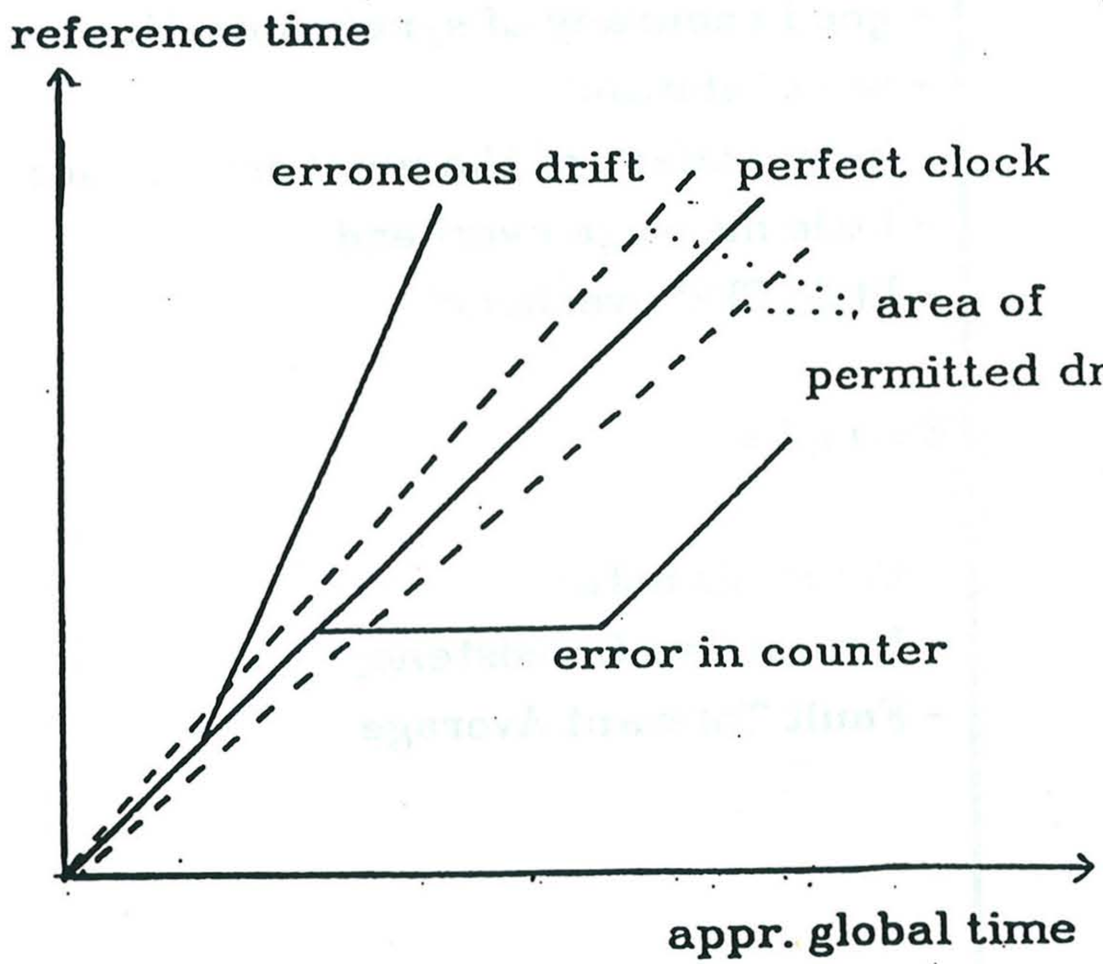
- good accuracy of synchronization
- fault tolerant
- independent of the number of nodes
- little message overhead
- little CPU overhead

Examples:

- Central Master
- Interactive Consistency
- Fault Tolerant Average

TU Wien		rt.8 R.10

Failure modes of a real time clock



Convergence Function

gives the maximum (worst case) difference of all good clock values immediately after instantaneous synchronization:

$$\Pi(\Delta^{int}, N, k, \varepsilon) = \Pi(\Delta^{int}, N, k) + \Pi(\varepsilon)$$

Drift rate

$$\xi = 2 \cdot \rho^{l} \cdot R_{int}$$

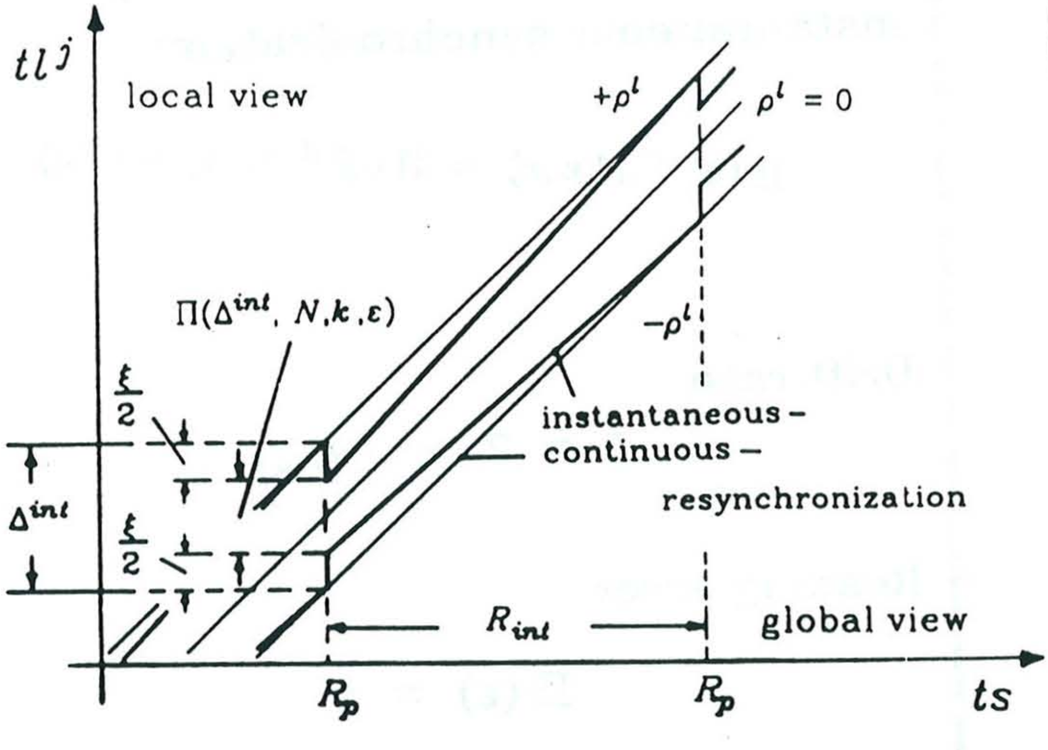
Reading **error**

$$\Pi(\varepsilon) = \varepsilon$$

Message loss or Byzanthine Fault:

$$\Pi(\Delta^{int}, N, k) = \frac{\Delta^{int}}{N - 2k}$$

Synchronization condition



Sychronization Condition:

$$\Pi + \xi = \Delta$$

Introducing a "divergence factor" d which is characteristic for the algorithm under investigation we get

$$(d \cdot \Delta + \varepsilon) + \xi = \Delta$$

which can be transformed to

$$\Delta = (\varepsilon + \xi) * 1/(1-d)$$

In the optimal case d = 0

If d=1 no synchronization is possible.

Synchronization Condition

$$\frac{\Pi(\Delta^{int}, \varepsilon)}{\Delta^{int} - \xi} \leq 1$$

Internal Synchronization Accuracy

$$\frac{\Delta^{int}}{\varepsilon + \xi} = \frac{N - 2k}{N - 3k} = u(N, k)$$

Faults k	Number of nodes N									
	4	5	6	7	8	9	10	15	20	30
1	2	1.5	1.33	1.25	1.2	1.16	1.14	1.08	1.06	1.03
2				3	2	1.66	1.5	1.22	1.14	1.08
3							4	1.5	1.27	1.14
4								2.33	1.5	1.22

Hardware Clock Synchronization in a system with 4 clocks (FTMP):

Trigger your clock to the second of the three other clocks.

Let us assume a Byzanthine clock D

A B C ----> time

then if	A	B	C
D early	B	(A)	(A)
D late	(C)	C	B

So after the synchronization, the clocks did not converge: d = 1

Let us now compare the FTA and the FTM algorithm for clock synchronization:

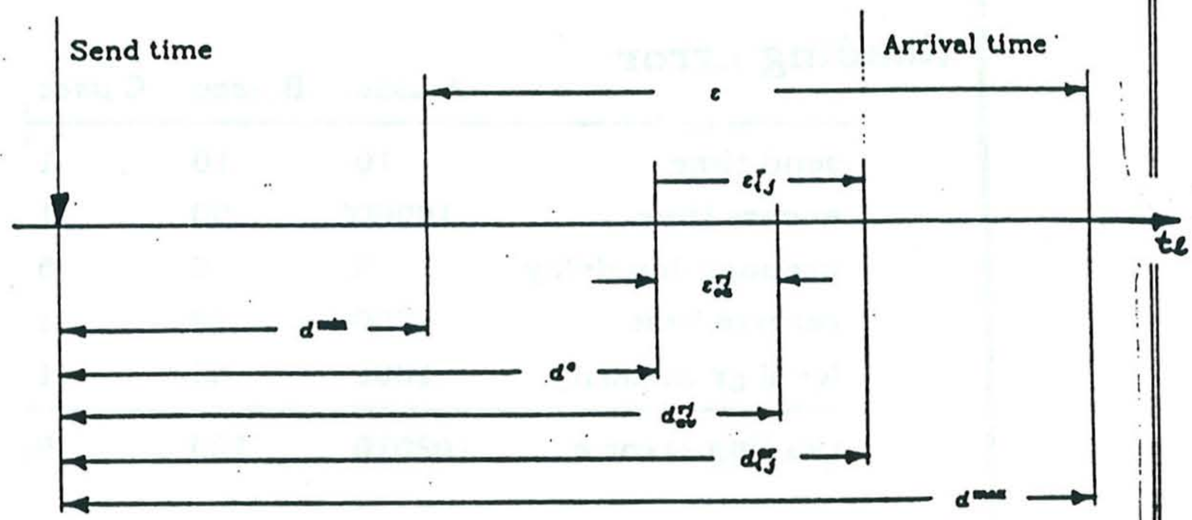
FTA d = k/(N-2k)

FTM d = 1/2

k=1	4	5	6
FTA	1/2	1/3	1/4
FTM	1/2	1/2	1/2

k=2	4	5	6
FTA	2/3	1/2	2/5
FTM	1/2	1/2	1/2

TU Wien		rt.9 R.11



r .. resynchronization period counter

d^r_{ij} .. message transfer delay from node i to node j in period r

ε^r_{ij} .. reading delay, i.e. the delay of message sent from node i to node j in period r

d^{min} .. minimum delay

d^e .. expected delay

d^{max} .. maximum delay

d^r_{aj} .. average delay of all resynchronization messages of the given synchronization period at node j

ε^r_{aj} .. deviation of the average delay from the expected delay

ε .. $d^{max} - d^{min}$ is called the reading error

Accuracy of Internal Synchronization:

Reading error

	A μsec	B μsec	C μsec
send time	10	10	1
access time	100000	50	1
propagation delay	5	5	5
receive time	1000	10	1
local granularity	1000	50	1
reading error ε	**102015**	**125**	**9**

Resynchronization deviation

($\rho = 5 \cdot 10^{-6}$)

resynchronization interval *sec*	1	10	100	1000
resynchronization deviation μsec	10	100	1000	10000

Total

$$(\varepsilon + \xi) \cdot u(N,k) = \Delta^{int} < 29\ \mu sec$$

External synchronization:

* Access to an external time reference (TAI via UTC)
* Whole ensemble is shifted to the external time
* Good long term stability, often low availablity

It must be the goal to provide a uniform timereference which is synchronized with the TAI.

This can be achieved with reasonable afford with a skew of about 100 microsecond

Reading error:

(1) variable time required to assemble and send the message after the local clock of the sender has been read (send time)

(2) variable medium access time (buffer)

(3) variable propagation delay (can be corrected in case of a single level LAN).

(4) variable time required to check the message and record the time of arrival (recive time)

(5) granularity of the local time

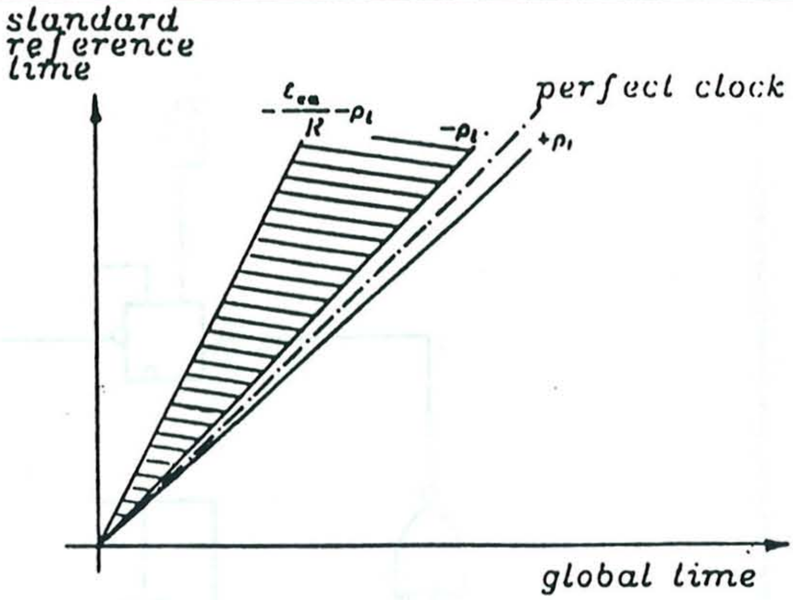


Fig. 4.2 Minimum delay correction ($d^* = d_{min}$)

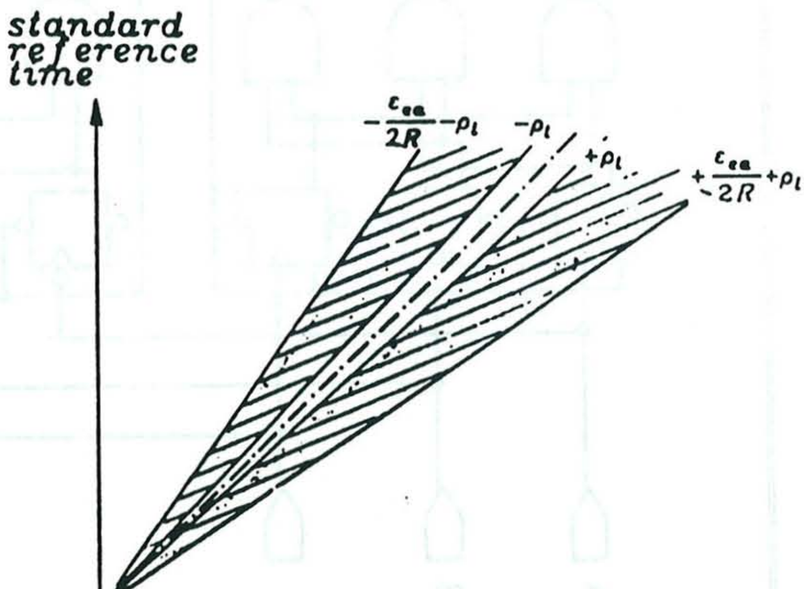
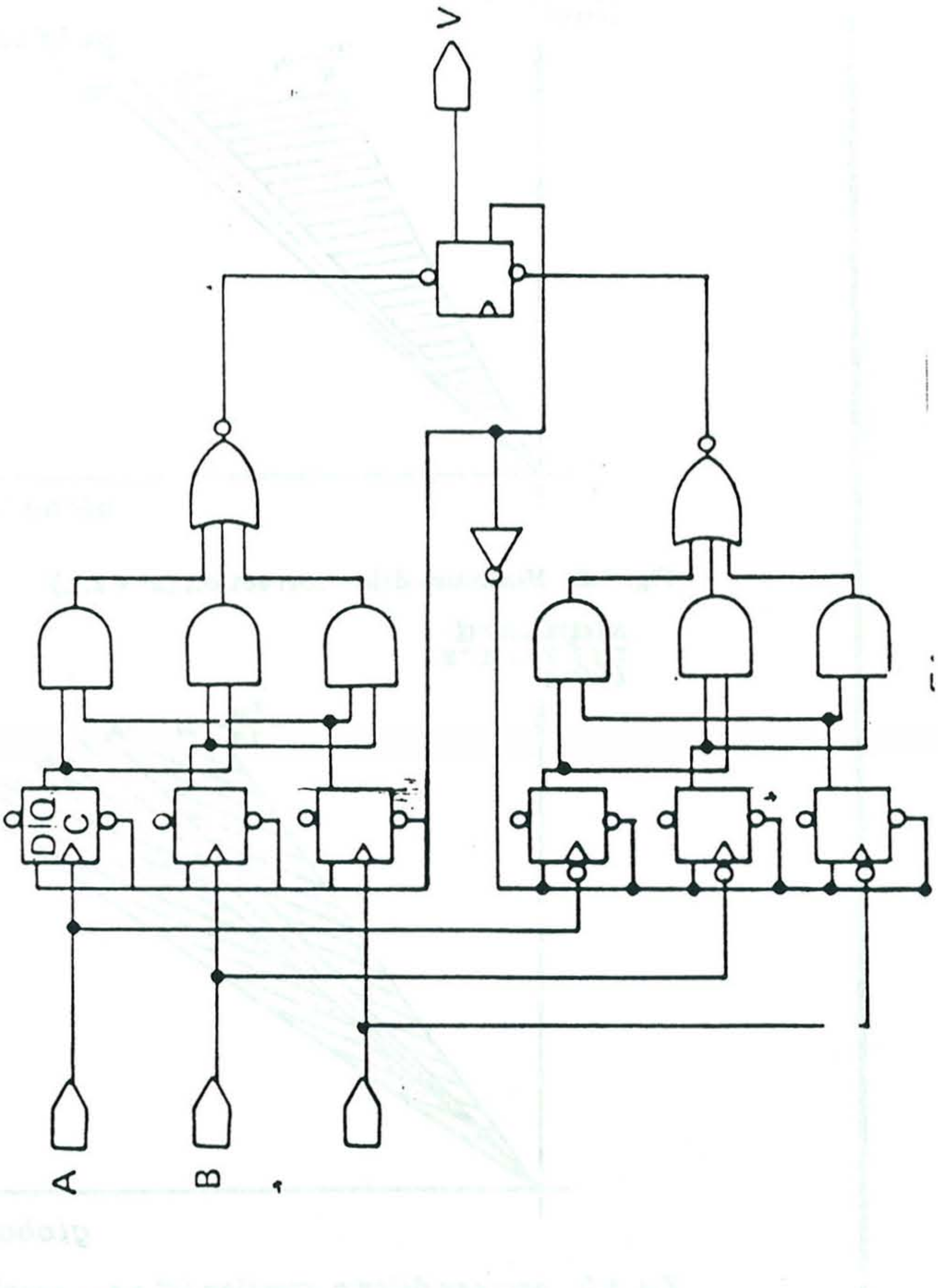
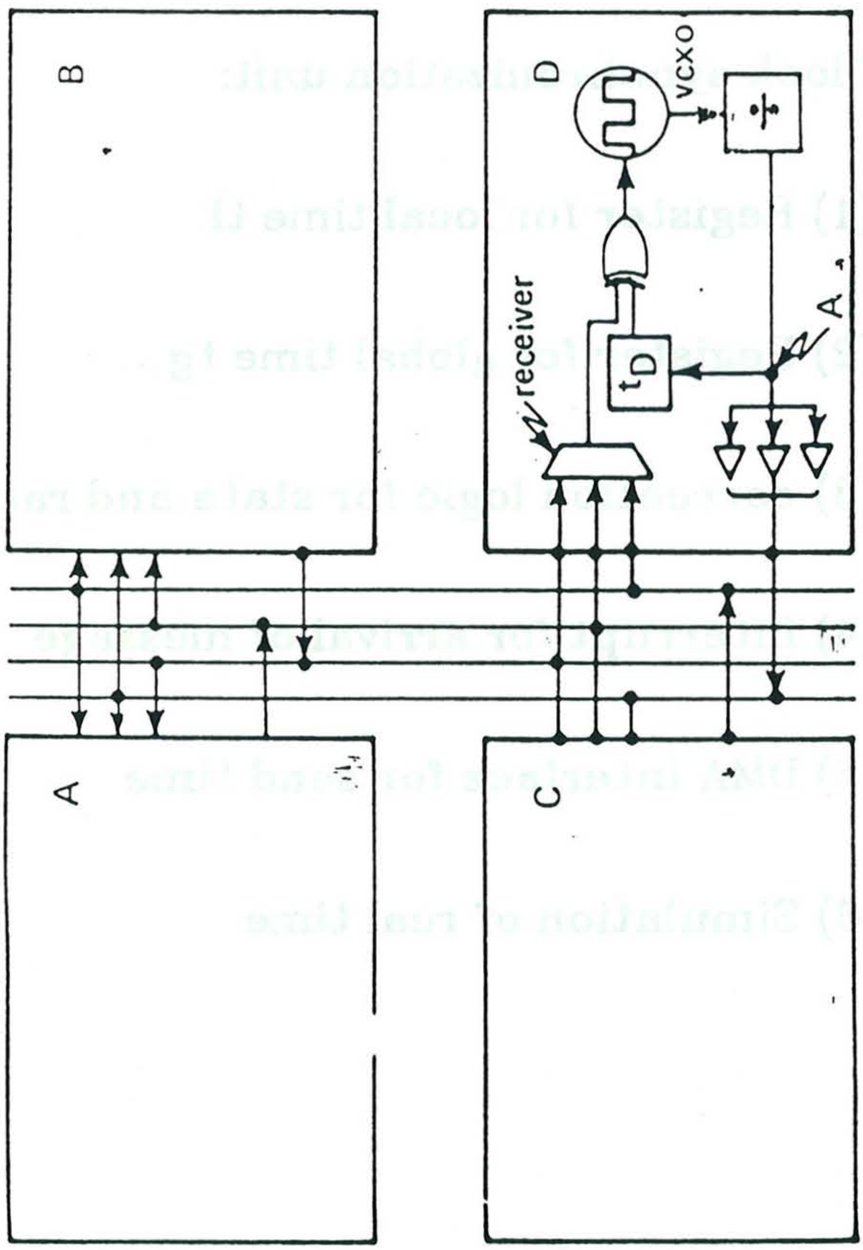


Fig. 4.3 Average delay correction ($d^* = d_{min} + \epsilon/2$)


V
D Q
C
A
B


B
D
VCXO
receiver
t_D
A
A
C

Clock synchronization unit:

(1) Register for local time tl

(2) Register for global time tg .

(3) correction logic for state and rate

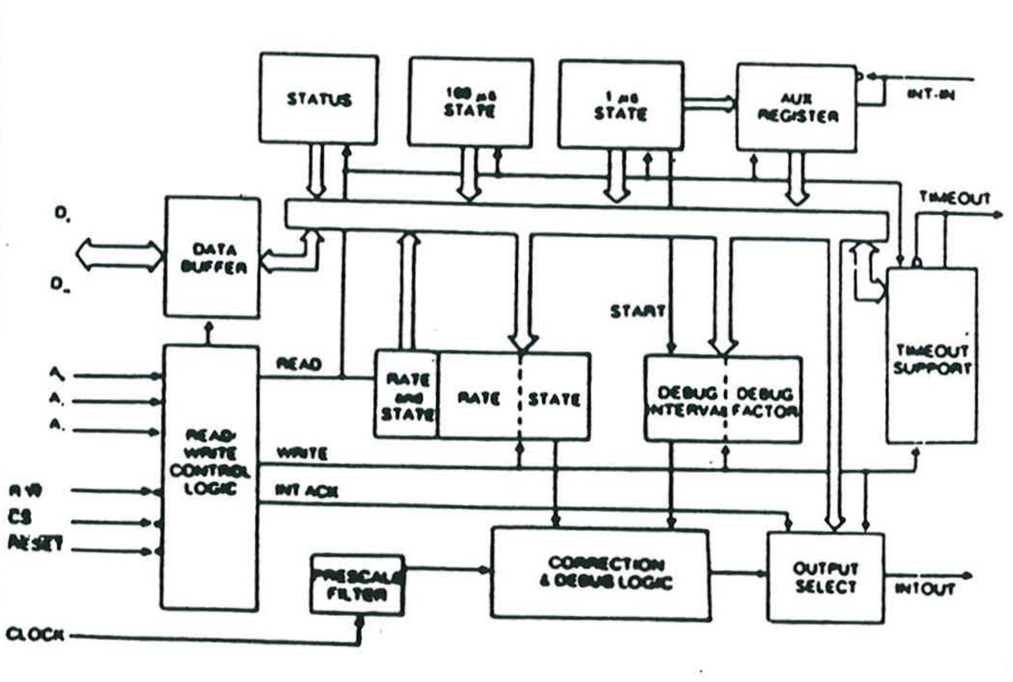
(4) Interrupt for arrival of message

(5) DMA Interface for send time

(6) Simulation of real time

TU Wien		vt. 19 R. 21

Block Diagram



DISCUSSION

Rapporteur: M.J. Elphick

Professor Wheeler asked if Professor Kopetz had considered the use of multiple observations, giving the possibility of increased accuracy by averaging. The speaker agreed that this would be possible, but was not used here. Following comments by Professor Kopetz on the best achievable accuracies (including the use of phase-locking hardware), Professor Wheeler queried the adjustment of clock values to take account of delays The speaker said that such second-order effects were disregarded; and in reply to a comment from Professor Randell, indicated that the clock *rates* were corrected at every re-synchronization cycle, to avoid discontinuities. In addition, every cluster was provided with a receiver for an external time standard. The synchronization circuits were digital, operating at 100mhz .

Another comment from Professor Randell asked about the approach the speaker would recommend for a wide-area distributed system, perhaps covering an entire country. Professor Kopetz answered that he would favour a combination of techniques, using a radio-frequency time signal between sites; this could provide both high availability and long-term stability. With the use of satellites, extremely accurate timing was possible, but problems of delay in transmission and distribution within the satellite had to be dealt with.