CACHE FEATURES

ARCHITECTURE
- SEPARATE I & D
- PHYSICAL CACHE
- DIRECT MAPPED
- WRITE THROUGH

ON-CHIP CACHE CONTROL
- BLOCK REFILL (1, 4, 8, 16, 32 WORDS)
- INSTRUCTION STREAMING
- DIRECT DRIVE OF SRAMS

IMPLEMENTATION
- USES STANDARD SRAMS
- WIDE RANGE OF SIZES & SPEEDS
- WRITE BUFFERS

PERFORMANCE ON M20000 SYSTEM:
- MISS RATE .005 FOR BLOCK of 16
  (FOR MIPS BENCHMARK)

CACHE ISSUES

PHYSICAL OR VIRTUAL

DIRECT MAPPED OR SET ASSOCIATIVE

STANDARD OR SPECIAL SRAMS

BLOCK REFILL SIZES AND STREAMING

ADVANTAGE OF PHYSICAL CACHE

- NO SYNONYMS
- NO PAGE COLORING
- NO REVERSE TRANSLATIONS
- REDUCES CACHE FLUSHING
- NO CYCLE TIME PENALTY IN PIPELINE
    - 1/2 CYCLE ON-CHIP TRANSLATION
    - FULL CYCLE FOR CACHE ACCESS
- WELL-SUITED FOR MULTIPROCESSOR DESIGN

MINIMIZING CACHE EFFECTS

- PARTIAL WORD STORE
- BLOCK REFILL
- INSTRUCTION STREAMING

PARTIAL WORD STORE

- KEEPS CACHE ENTRY VALID

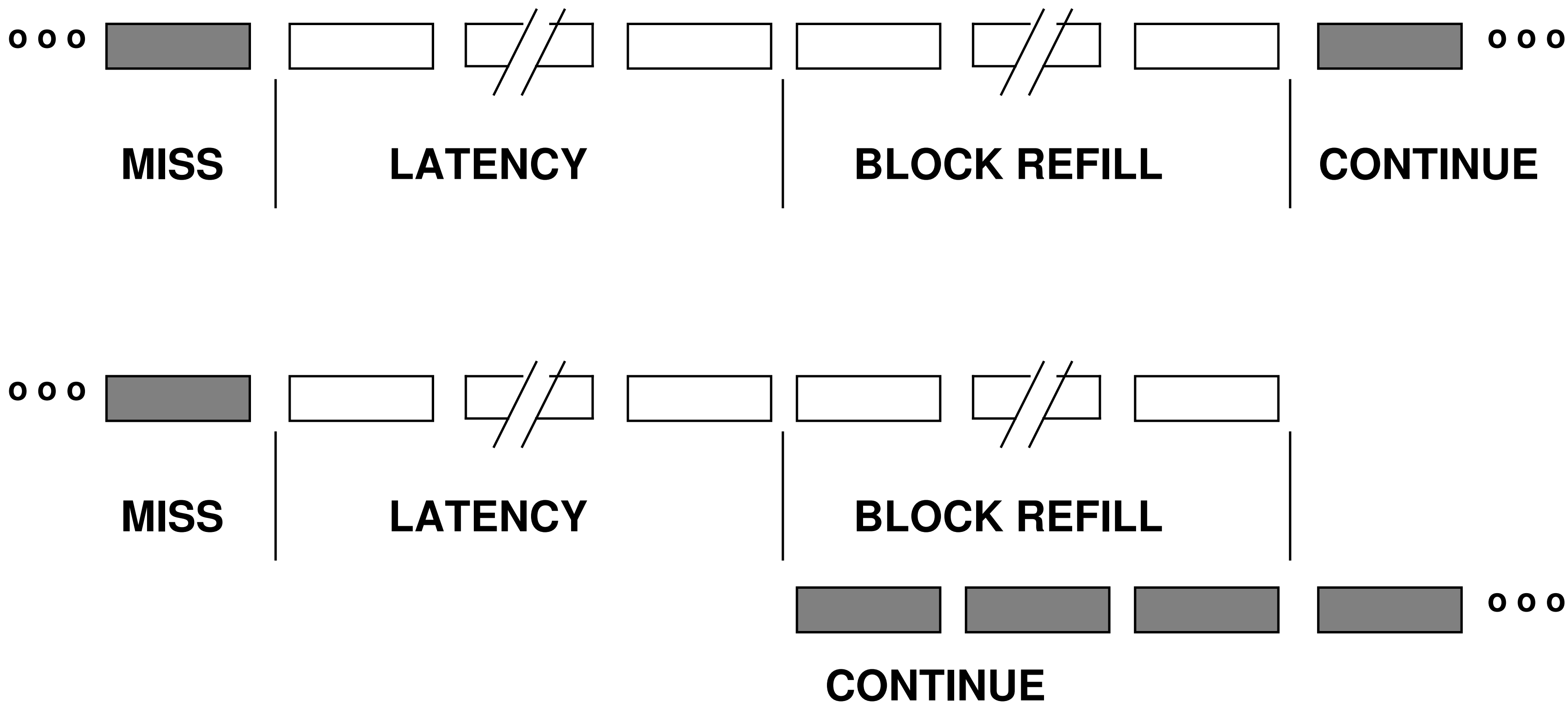
- PARTIAL WORD STORES USE READ

-MODIFY-WRITES- ELIMINATES SEPARATE WRITE CONTROLS

MIPS BLOCK REFILL

- AMORTIZE COST OF LATENCY OVER MULTIPLE WORDS
- CPU HARDWARE CONTROLS THE TRANSFER
- CHOICE OF 4, 8, 16, or 32 WORD BLOCK SIZES
- SEPARATE INSTRUCTION AND DATA BLOCK SIZES
- DESIRABLE TO TUNE BLOCK SIZE TO SYSTEM DESIGN, INCREASE BLOCK SIZE AS LATENCY TO MEMORY INCREASES
- BLOCK REFILL HAS LARGE IMPACT ON HIT RATE, SECOND ONLY TO TOTAL CACHE SIZE

MIPS INSTRUCTION STREAMING



MIPS INSTRUCTION STREAMING

- BEGIN EXECUTION WHEN INSTRUCTION ARRIVES, ISSUE INSTRUCTION IN SAME CYCLE AS CACHE WRITE

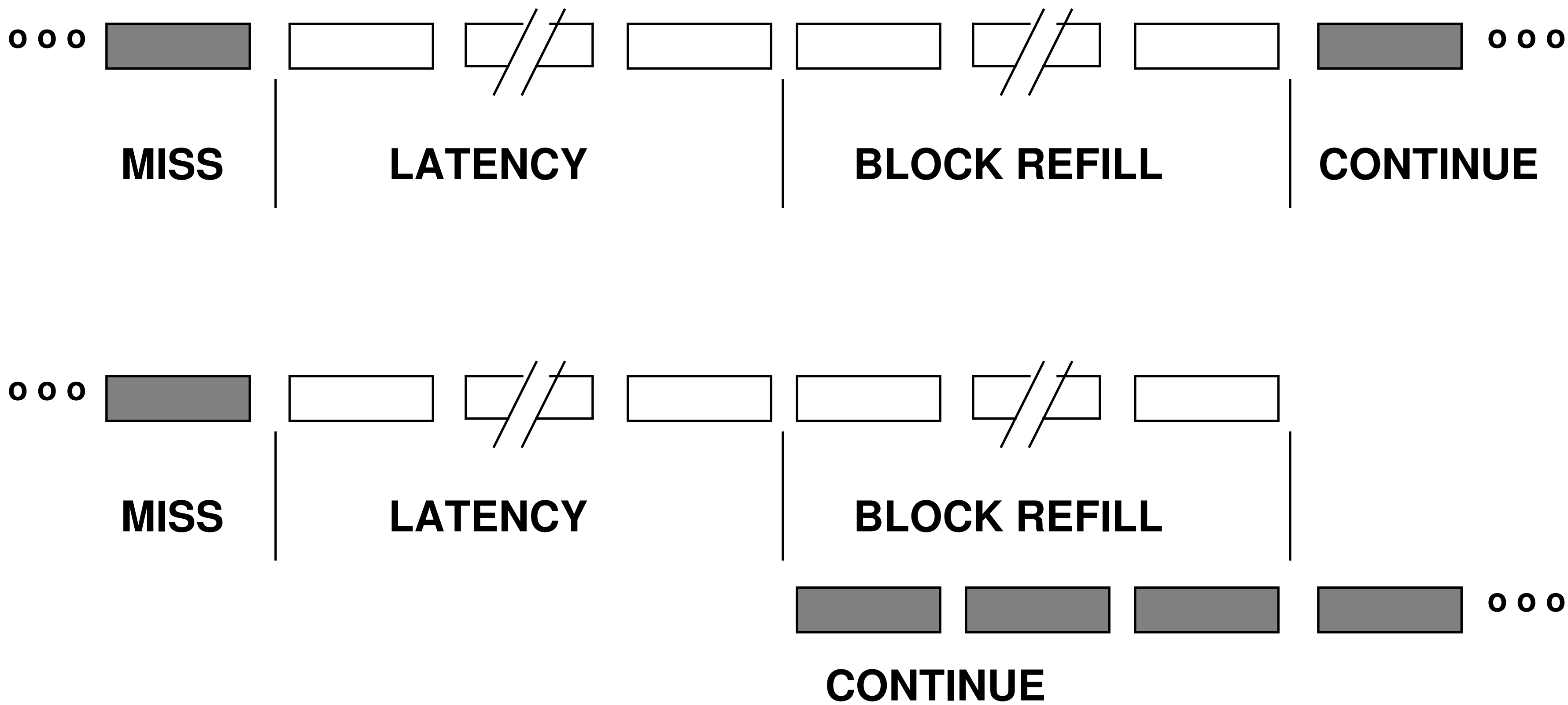
- 70% OF INSTRUCTION MISSES START AT BLOCK BOUNDRY, GREATEST BENEFIT TO STREAMING INSTRUCTIONS

- MAXIMUM EXECUTION SPEED MATCHES MAX MAIN MEMORY BANDWIDTH, 100MB/SEC

- 5% NET GAIN - vs SYSTEM WITHOUT INSTRUCTION STREAMING

LR3000 CACHE DESIGN

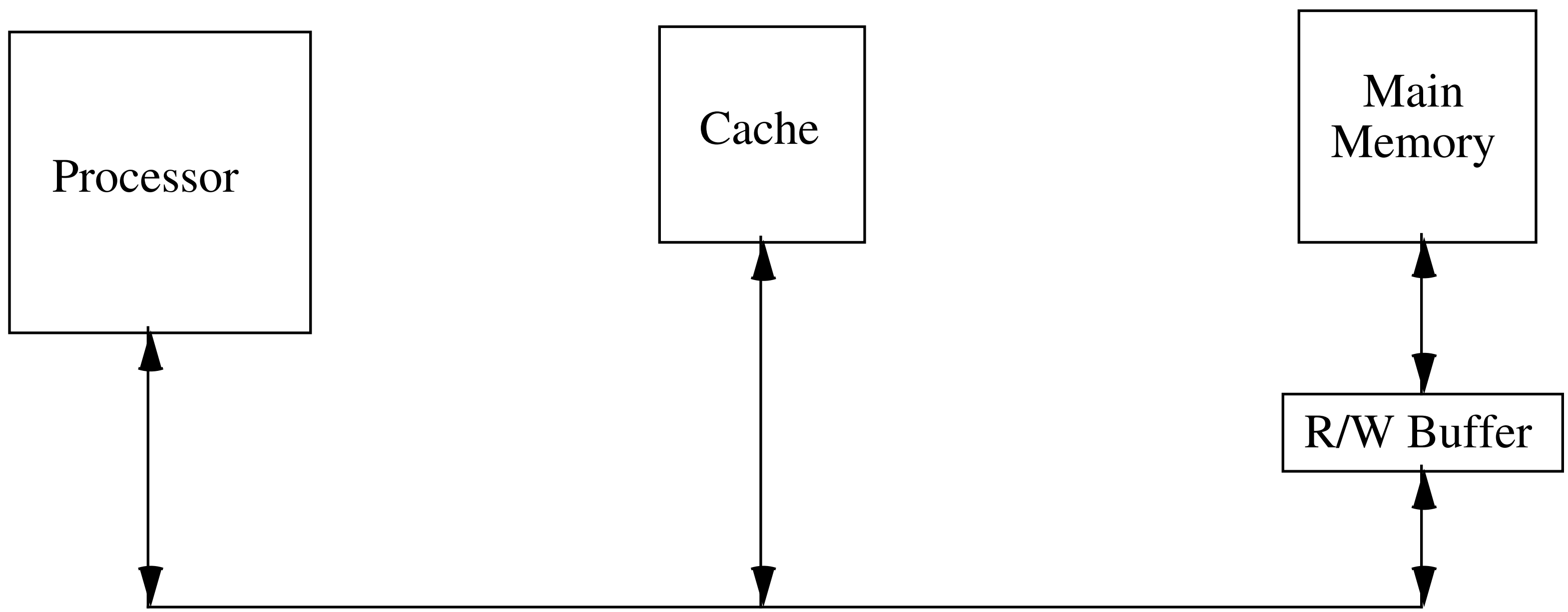
LR3000 CACHE DESIGN



LR3000 CACHE DESIGN

- Cache Introduction
- LR3000 Cache Interface
- Determining the Required Clock Delays
- Small Systems

WHAT IS A CACHE?

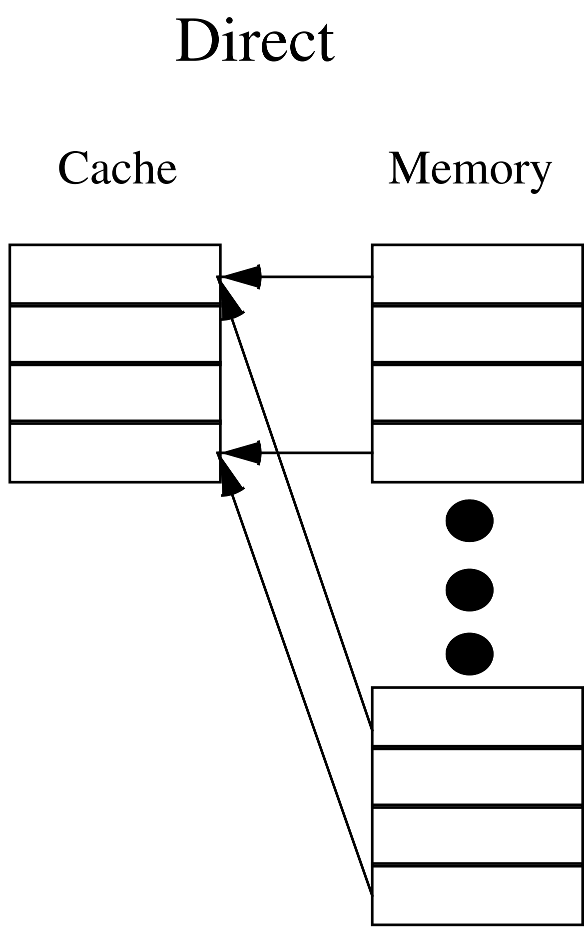


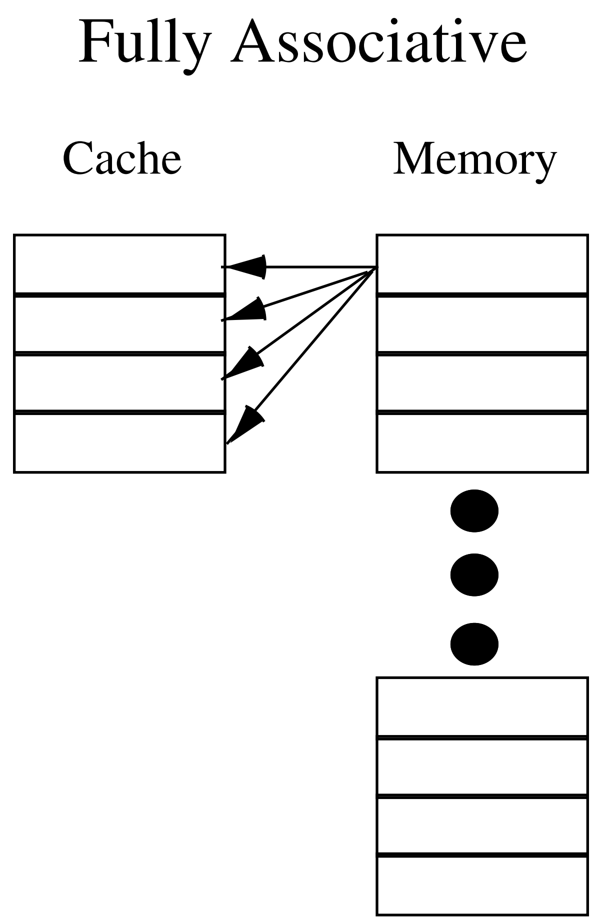
- Fast memory between the processor and the main memory

CONSIDERATIONS FOR A CACHED SYSTEM

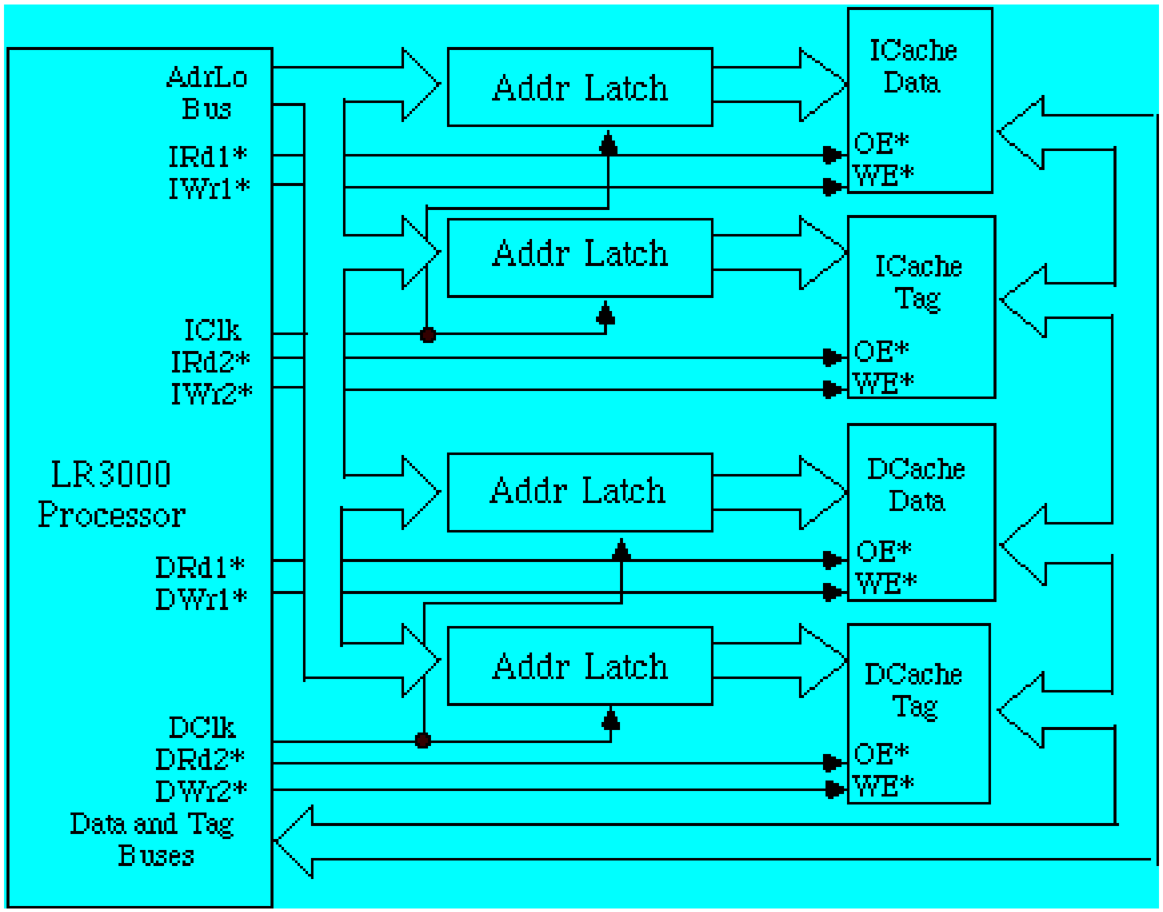
- Cacheable accesses with hits improve system performance
- Misses result in penalties
- Mapping strategies - Direct, Set associative, Fully associative
- Replacement strategies - LRU, Random, LFU, etc.
- Needs initialization

MAPPING STRATEGIES

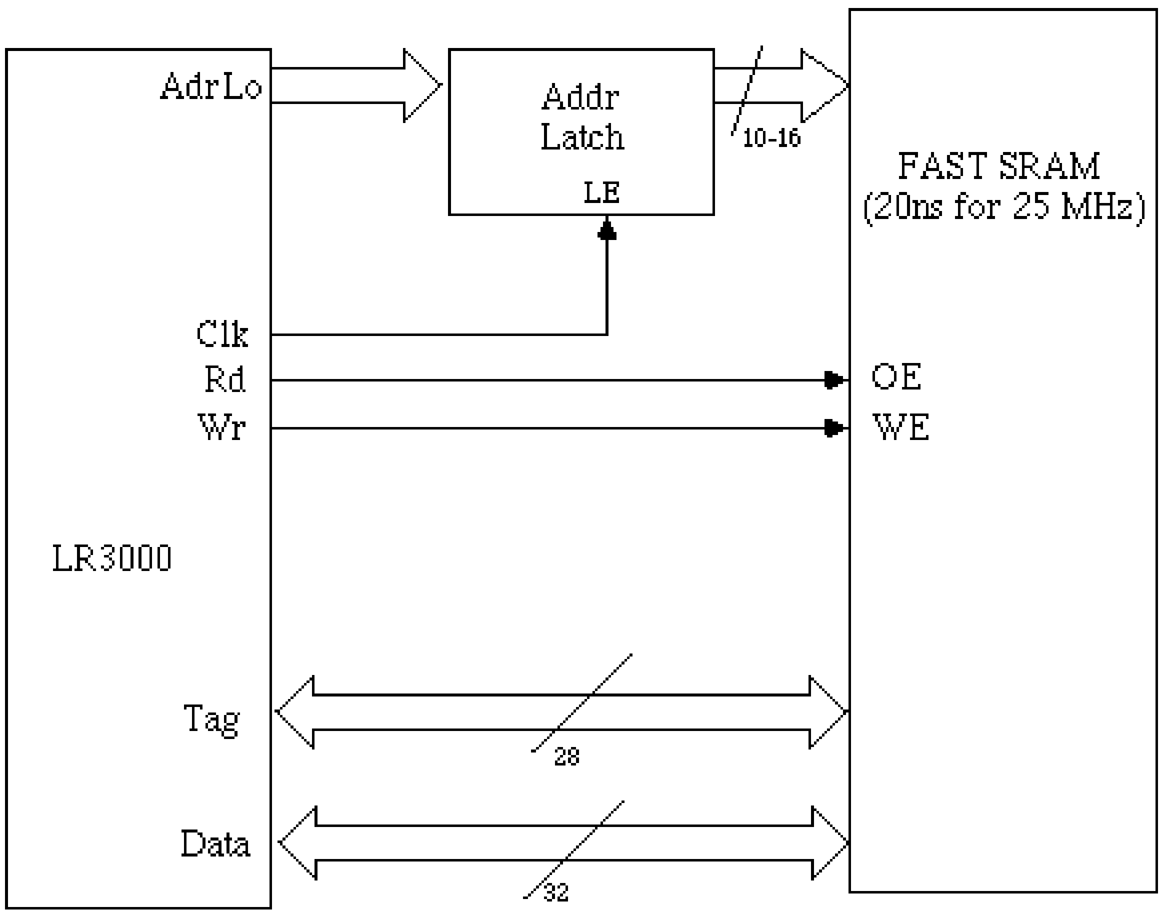




CACHE INTERFACE



CACHE INTERFACE



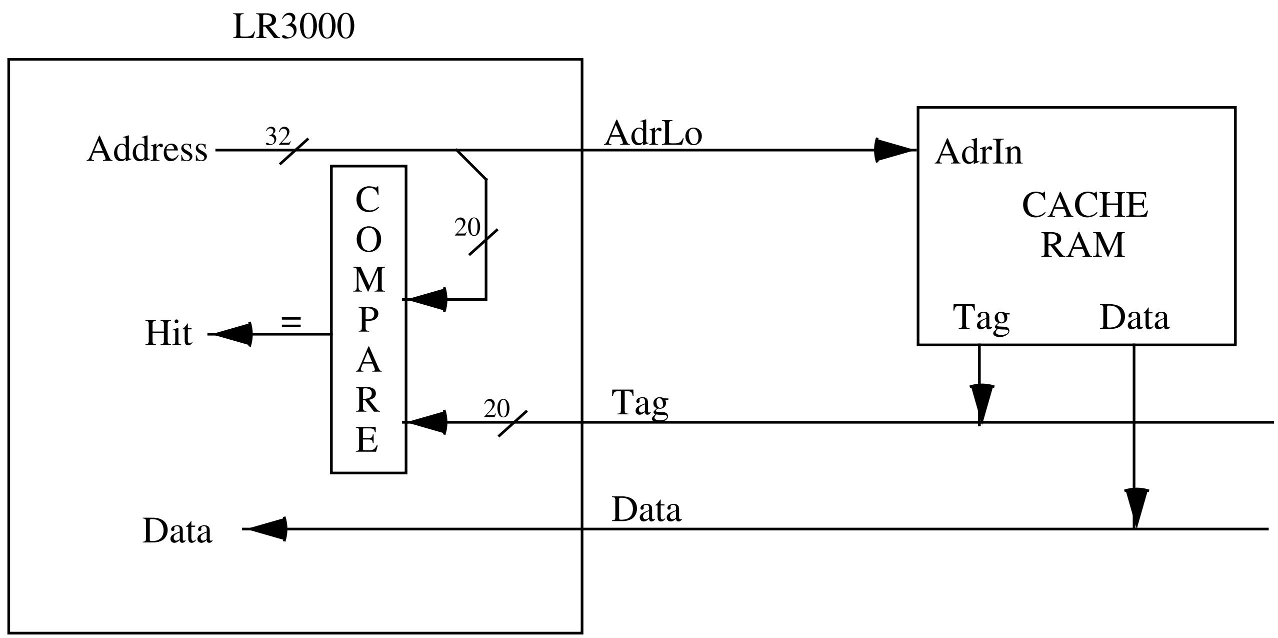
SUPPORTED CACHE SIZES

Bytes	Words	AddrSize	AdrLo Bits
4 K	1024	10	11:2
8 K	2048	11	12:2
16 K	4096	12	13:2
32 K	8192	13	14:2
64 K	16384	14	15:2
128 K	32768	15	16:2*
256 K	65536	16	17:2*

* Requires Extended Cache (ExCache) Mode and Uniprocessor Configuration

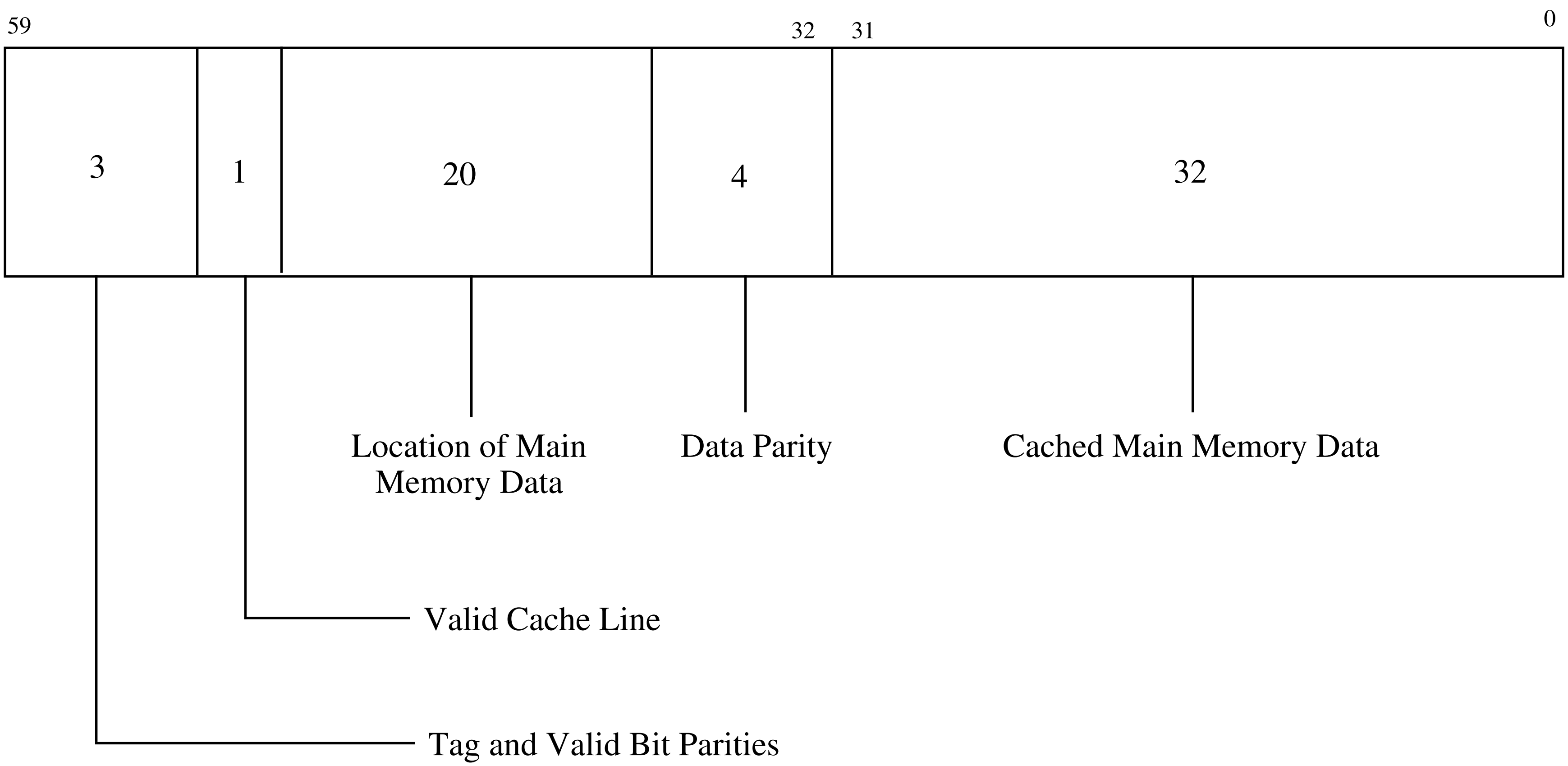
64KBytes is Assumed for this Class

CACHE LOOK-UP (SIMPLIFIED)



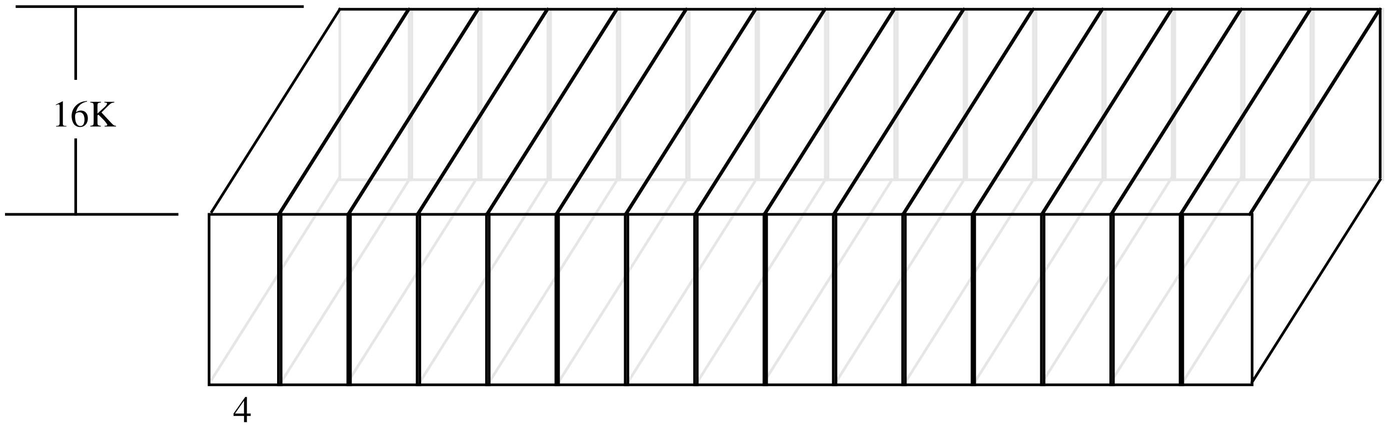
- A Hit Also Requires Both Parity Fields to be Valid and the Valid Bit to be Set

CACHE LINE FORMAT

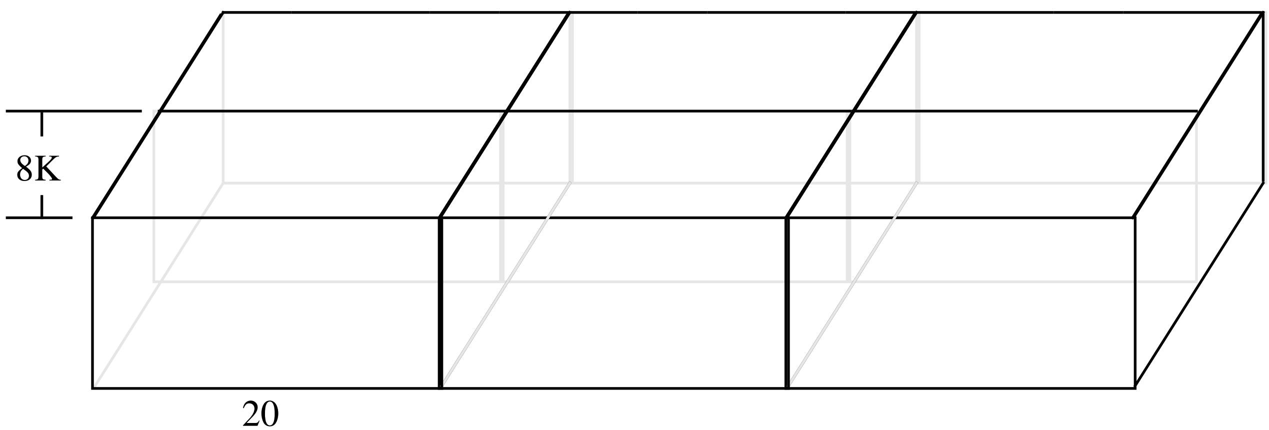


64KBYTE CACHE IMPLEMENTATIONS

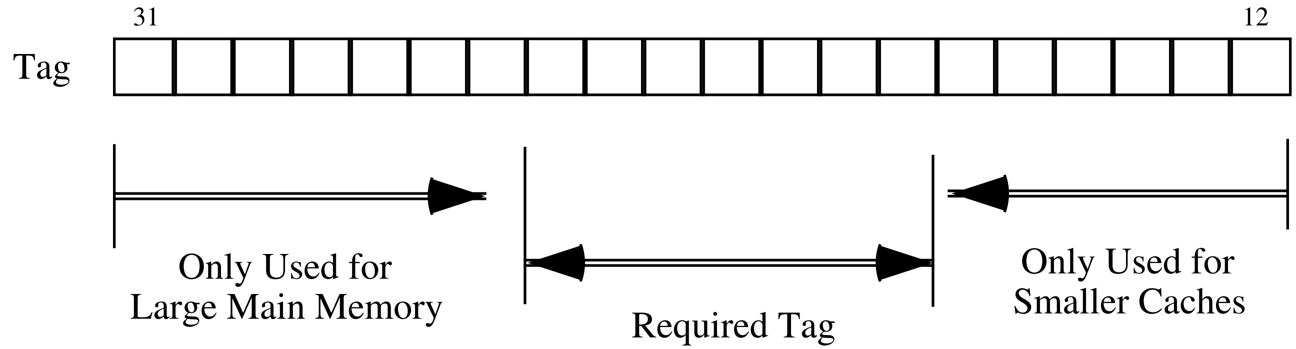
16Kx4 (15):



8Kx20 (6):



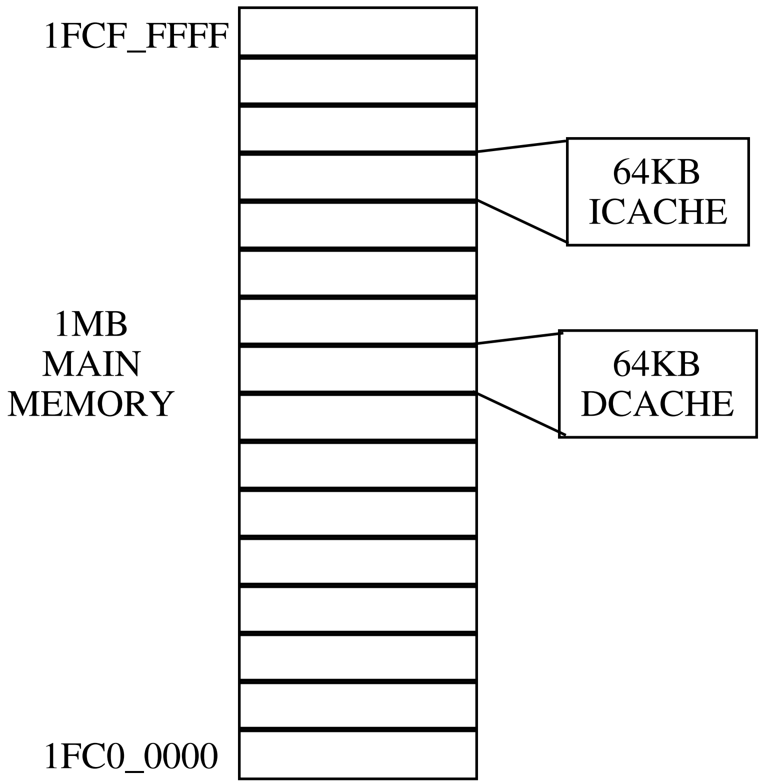
REDUNDANT TAG BITS

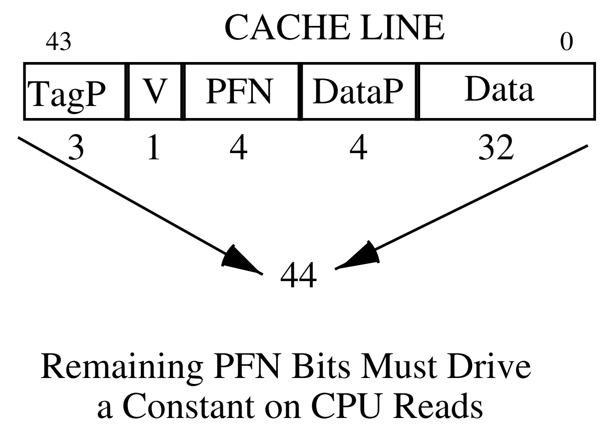


- Tag Identifies Where the Cache Word Came From in Main Memory
- Smaller Main Memory Requires Less Tag Bits
- Larger Cache Requires Less Tag Bits

SMALL SYSTEMS

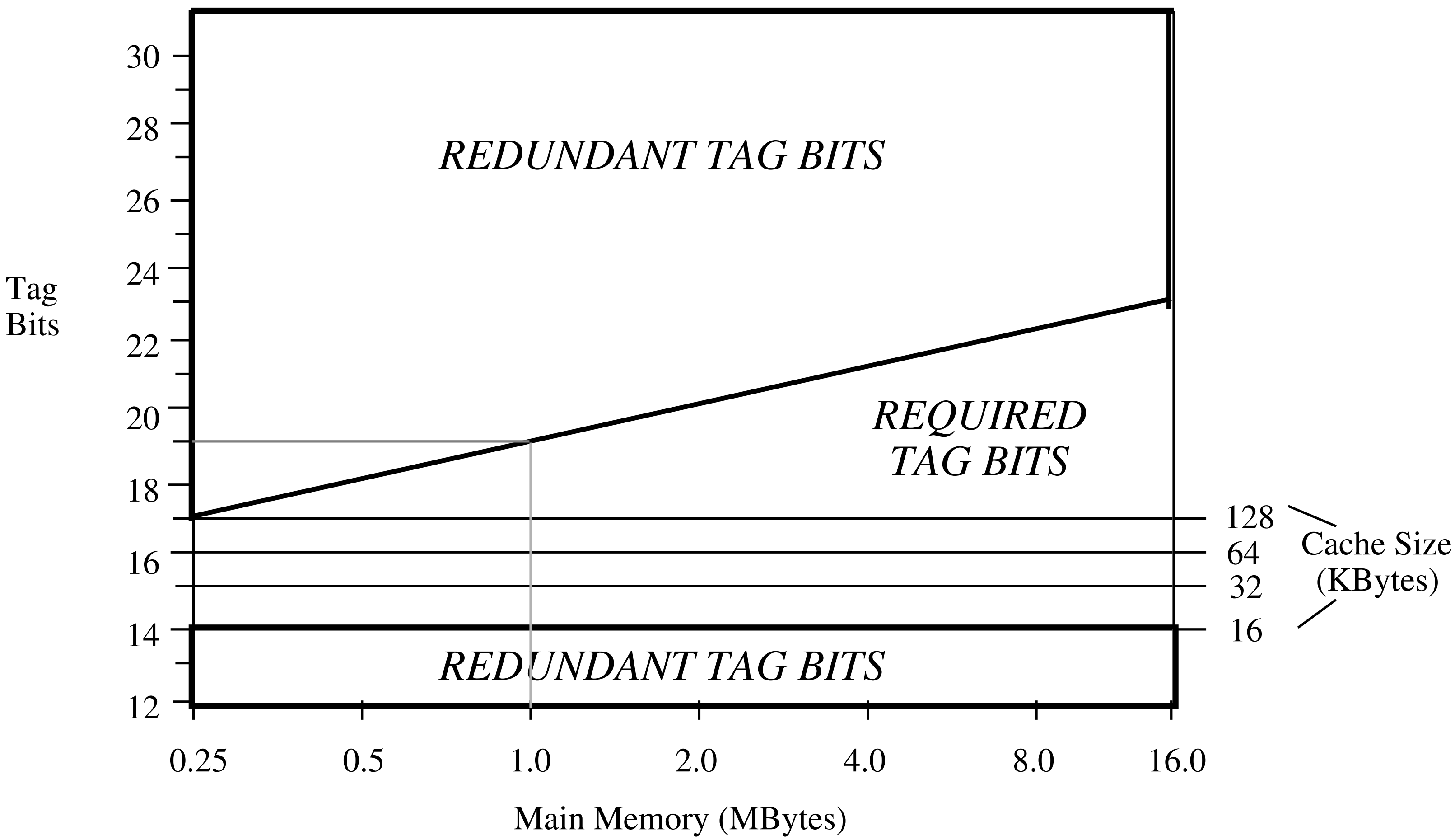
EXAMPLE: 1MByte of Physical Memory, 64KByte Caches





- Each Cache Line can Come From One Of 16 Places (Tag Size = 4)
- Bootstrap Vectors Start at 1FC0_0000

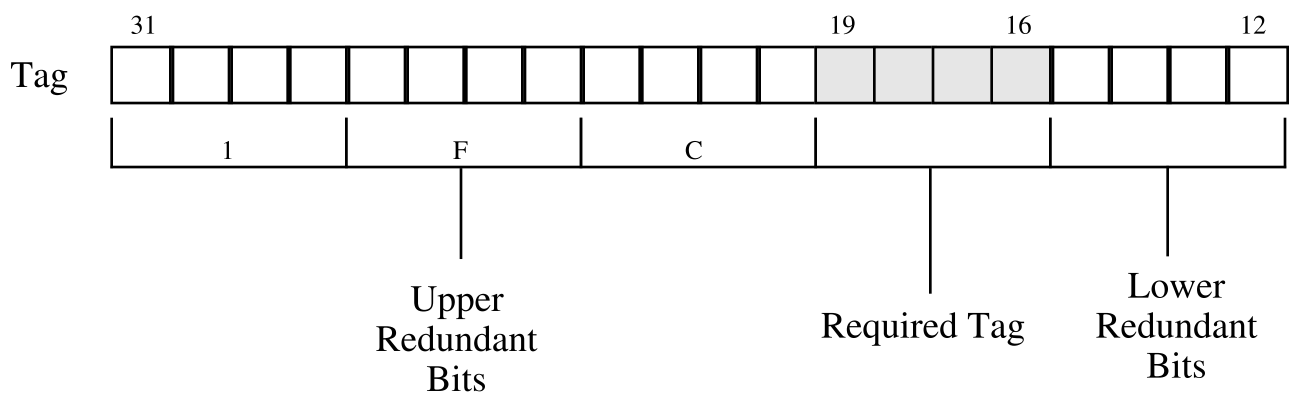
REDUNDANT TAG BITS CALCULATION



EXAMPLE: 1MB Main Memory & 64KB Caches; Tag = 16..19

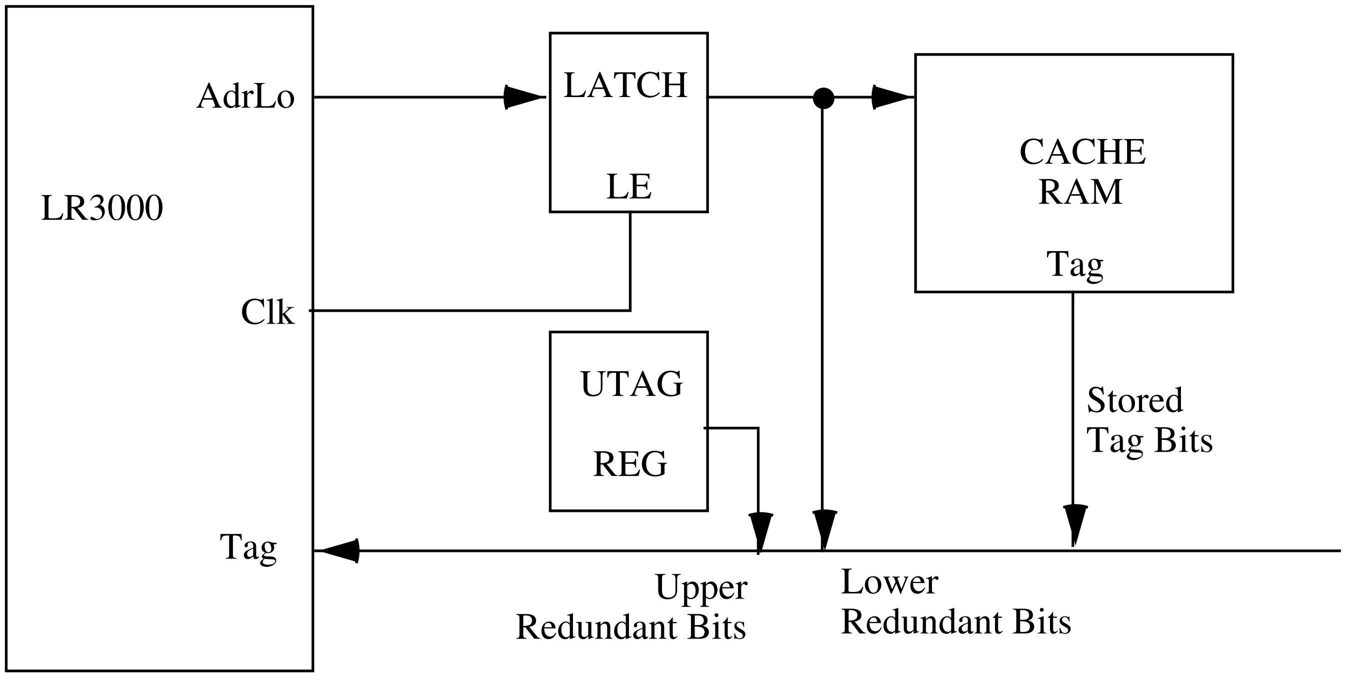
TAG ORGANIZATION (SMALL SYSTEM)

EXAMPLE: 1MByte of Physical Memory, 64KByte Caches



- Use FCT241A to Drive Tag Bus
- Save 4 RAM Packages (16K x 4) per Cache
- No RAM Package Savings for 8K x 20

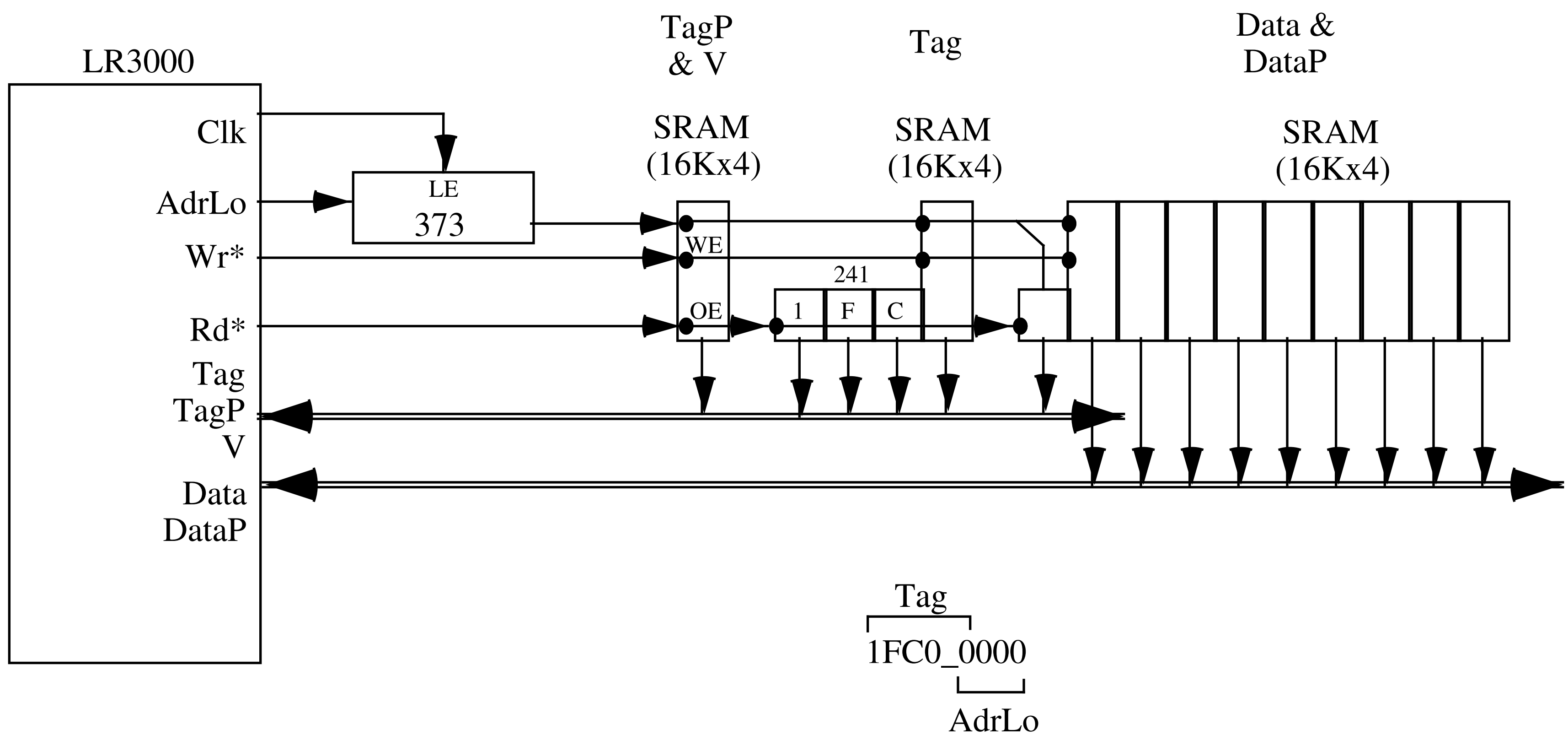
SUPPLYING THE REDUNDANT TAG BITS



- Processor Should Store Start Address of Main Memory in UTAG Register

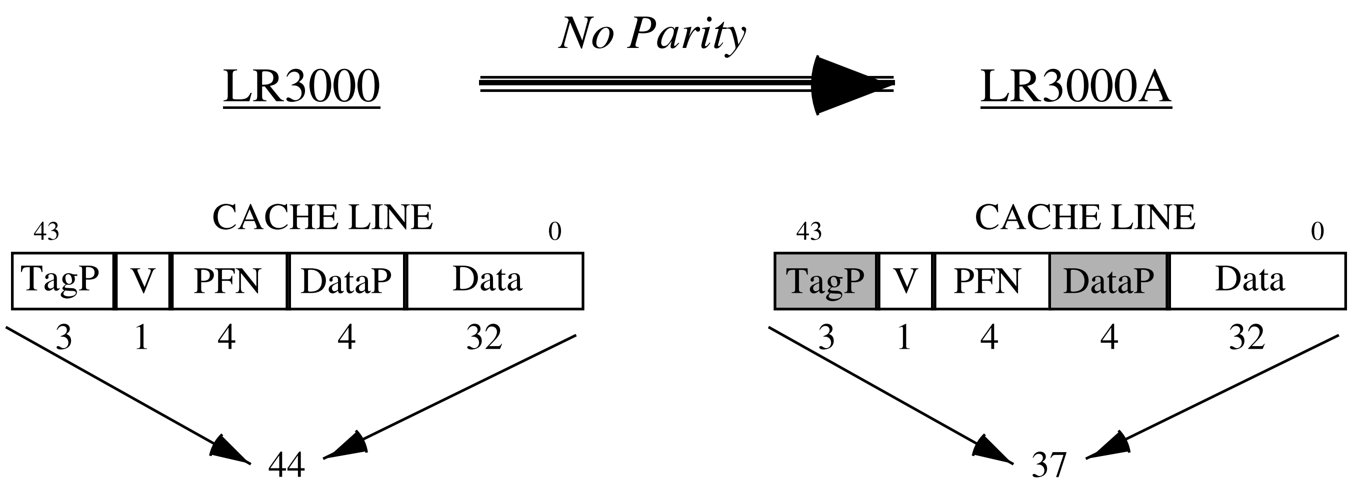
SMALL SYSTEMS

EXAMPLE: 1MByte of Physical Memory, 64KByte Caches



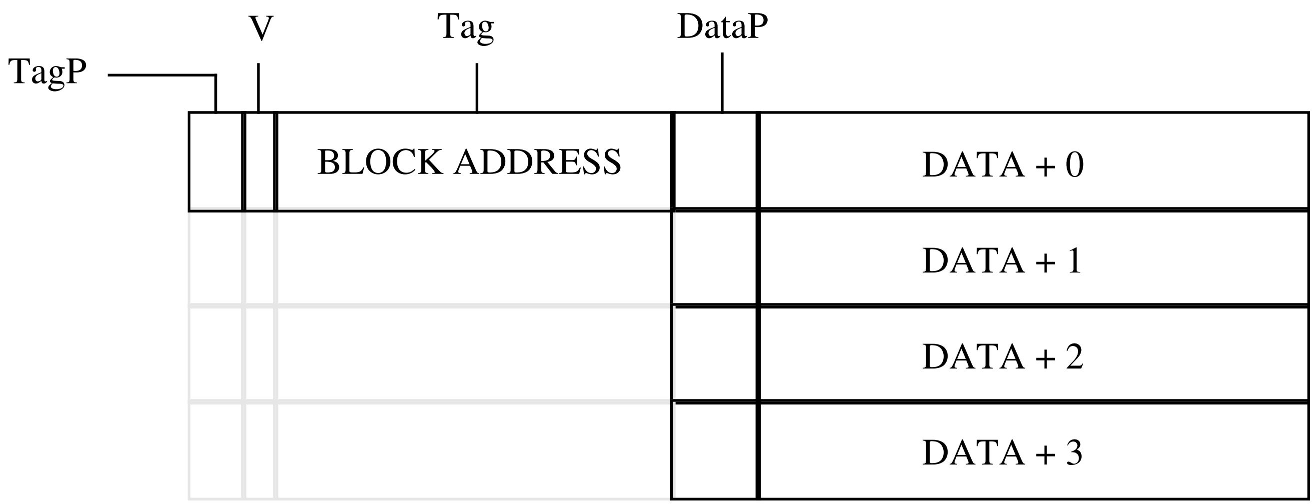
SMALL SYSTEMS

EXAMPLE: 1MByte of Physical Memory, 64KByte Caches
and No Parity Requirement



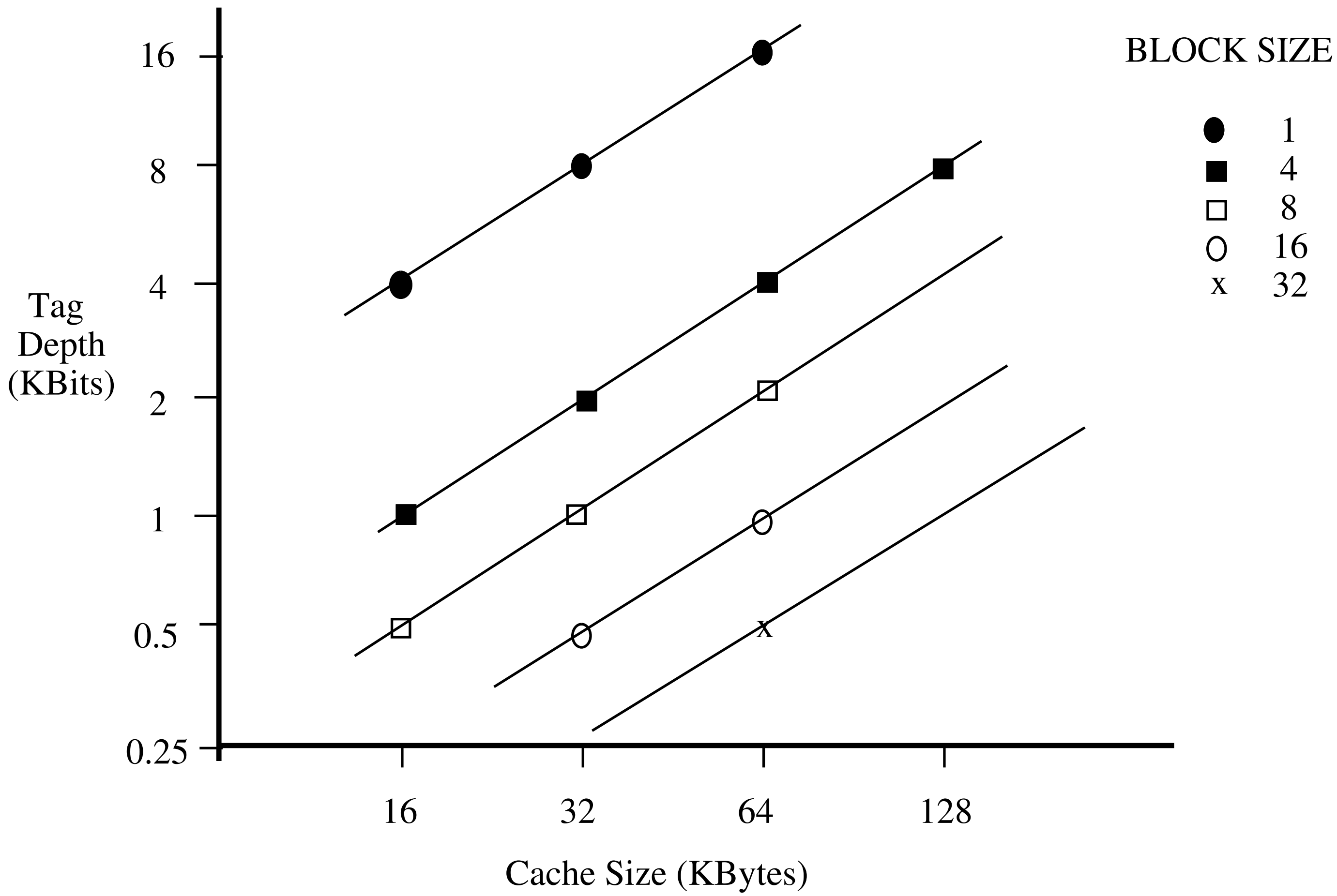
- LR3000A Allows Elimination of an Extra 16K x 4 Package or 2 8K x 20 Packages

REDUNDANT TAG ENTRIES



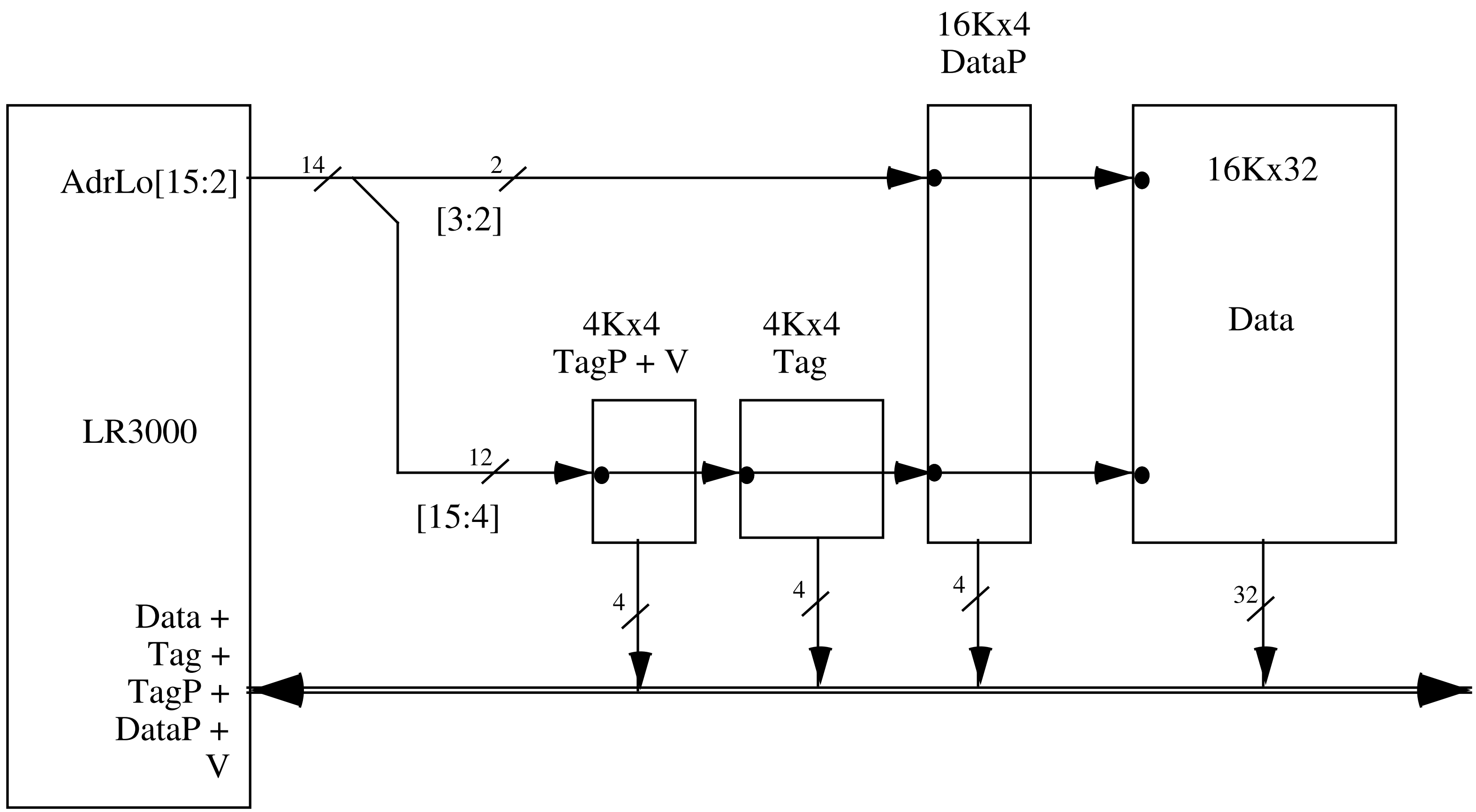
- Tag is Same for All Entries of a Block
- Tag Memory Depth can be Reduced

REDUNDANT TAG ENTRIES



REDUNDANT TAG ENTRIES

EXAMPLE: 1MByte Physical Memory, 64KByte Caches,
4 Word Block Size

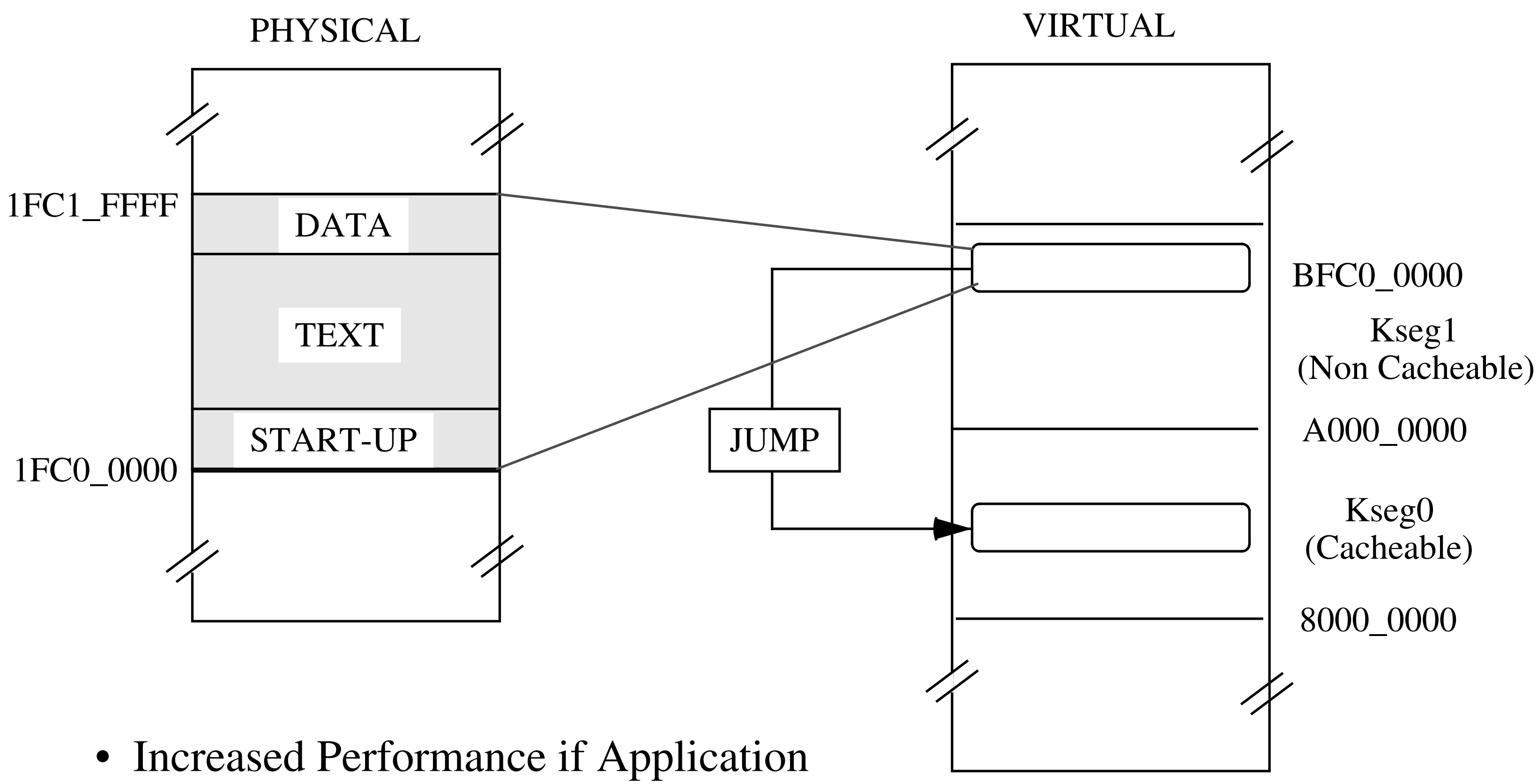


<u>**REDUNDANT TAG ENTRIES**</u>

<u>*Useful Only for ICache !!!*</u>

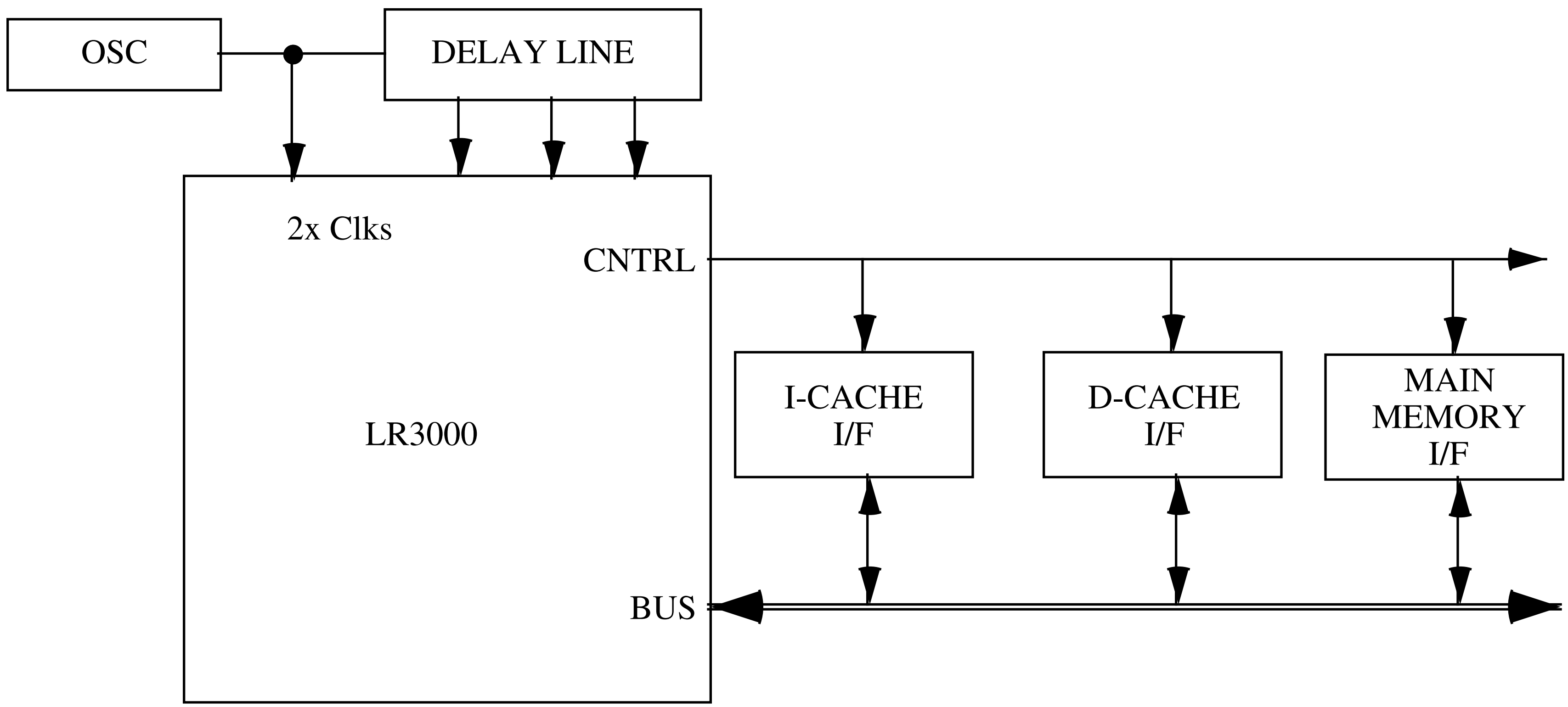
SMALL SYSTEMS

EXAMPLE: 128KByte of Physical Memory, 64KByte Caches



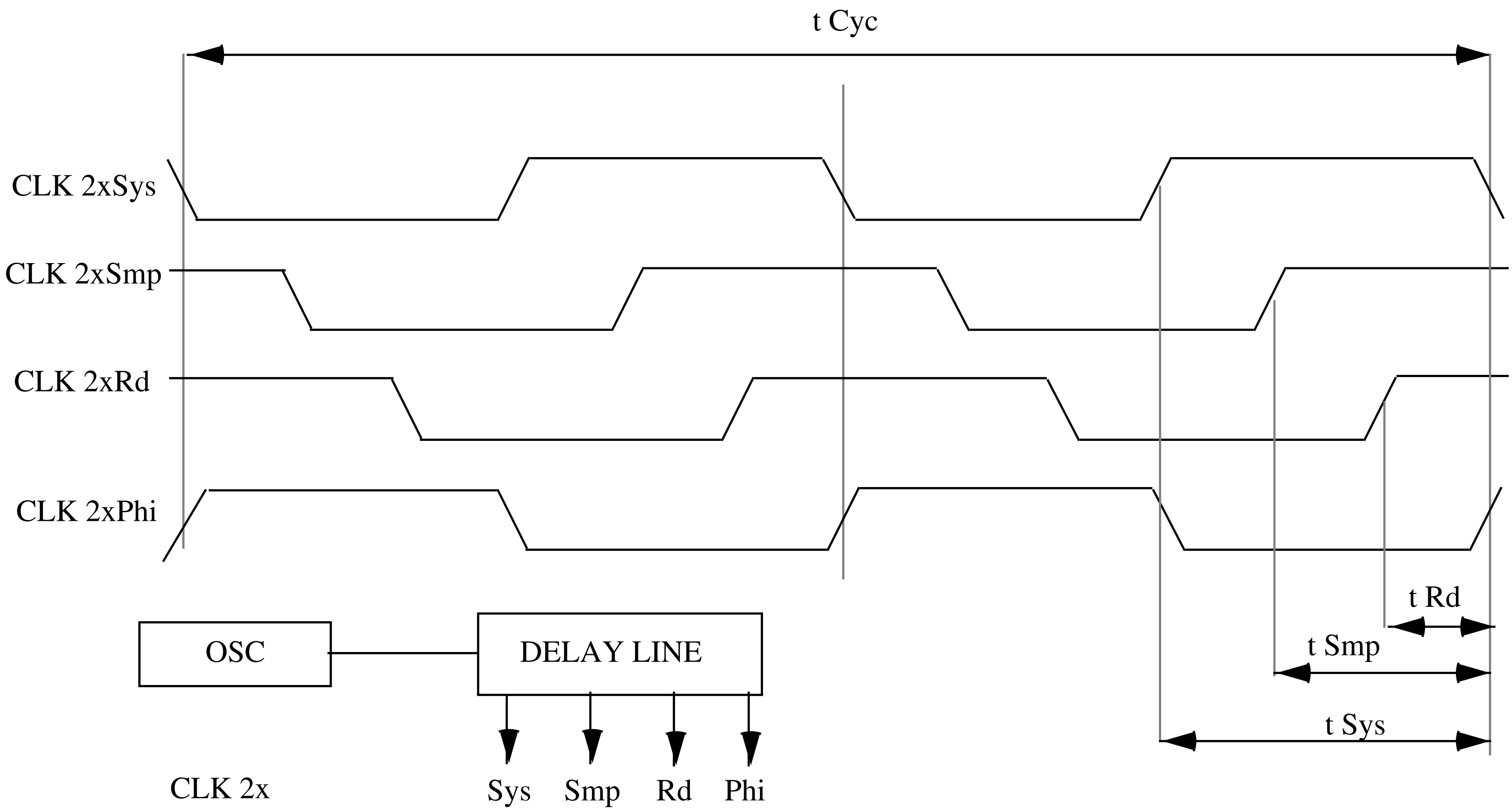
- Increased Performance if Application Can Fit in Cache

CACHE DESIGN



- Cache Design is Choosing Correct Clock Delays to Permit Reliable Data Transfer Among the Three Major System Components

THE FOUR PHASE CLOCK

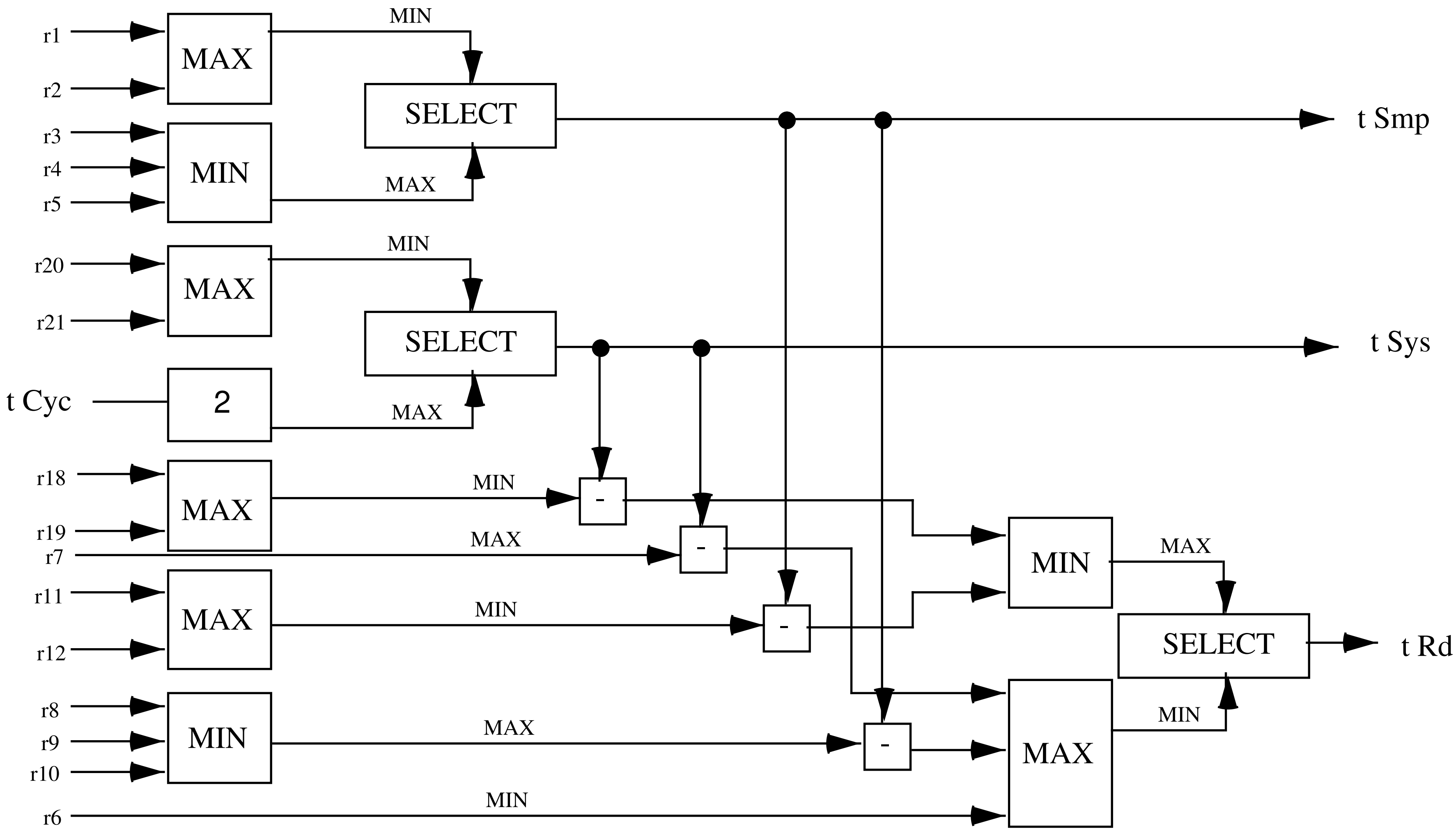


TWO DESIGN APPROACHES

1. Use "Riordan" Formulas to Guarantee Correct Timing for Each Parameter.

2. Make a Few Assumptions and use Simple 3 Step Process to Set Clock Delay Values

RIORDAN CLOCK DELAY CALCULATIONS

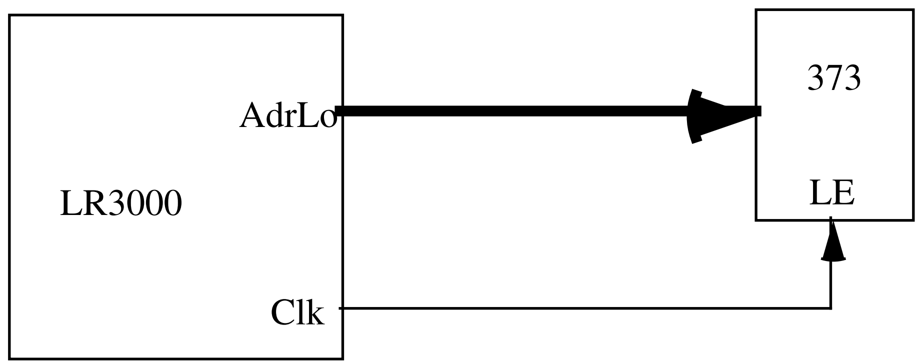


CALCULATING t Smp

FORMULA	DESCRIPTION	FOR 25MHz ns
r1	Internal Sample to Phase Delay	5
r2	Address Capture by IClk & DClk	1.8
r3	Address Access to Sample	6.8
r4	Address Setup to End of Write	19.8
r5	Processor Data Setup to End of Write	7

RIORDAN 2

Address Capture by IClk and DClk - This constraint guarantees that the transparent latches on the AdrLo bus close before the processor changes the address.



t Smp   t373Hld + tClkDeMin - tAdrLoDeMin

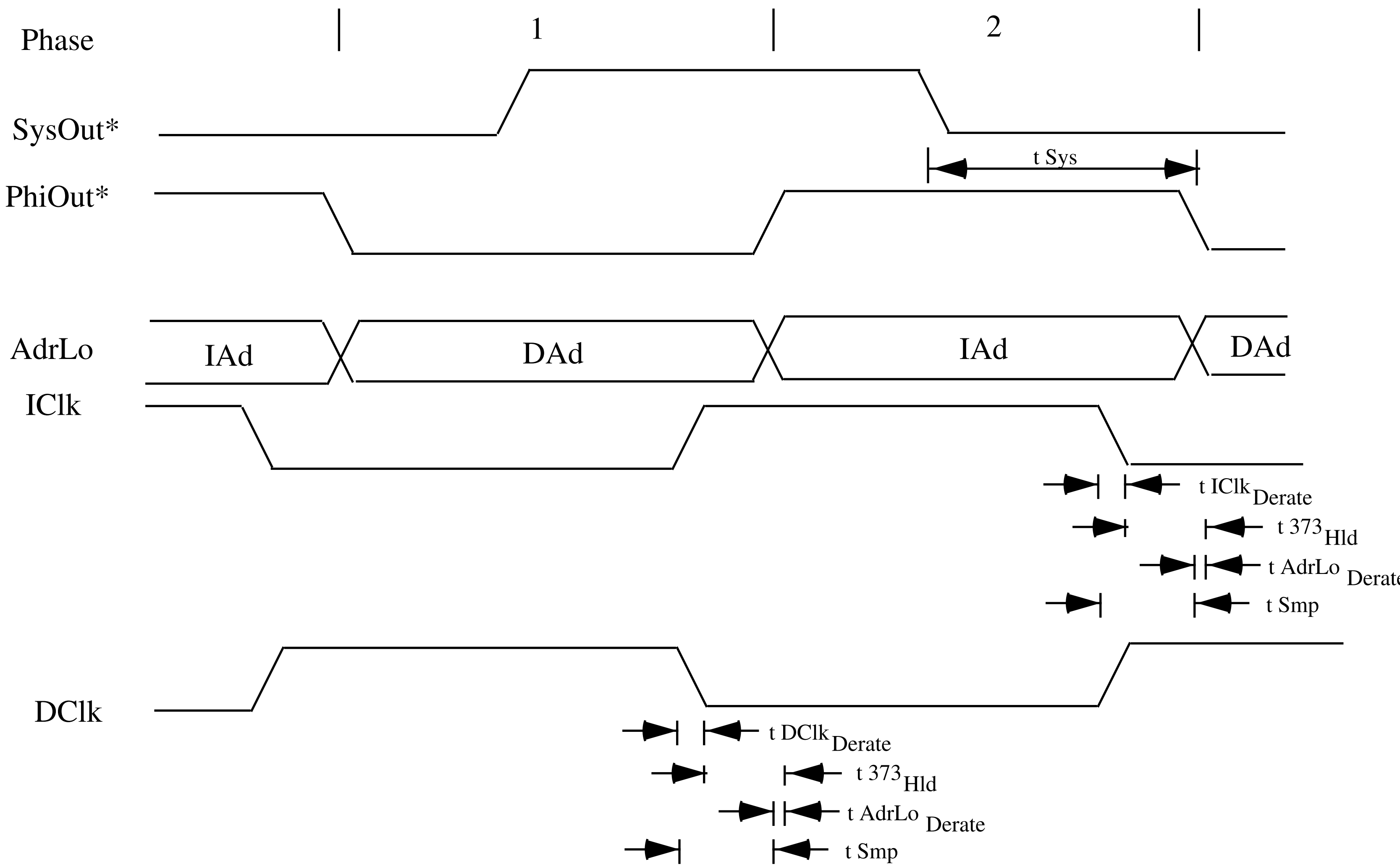
= 1.8 + 0.5 - 0.5

= 1.8

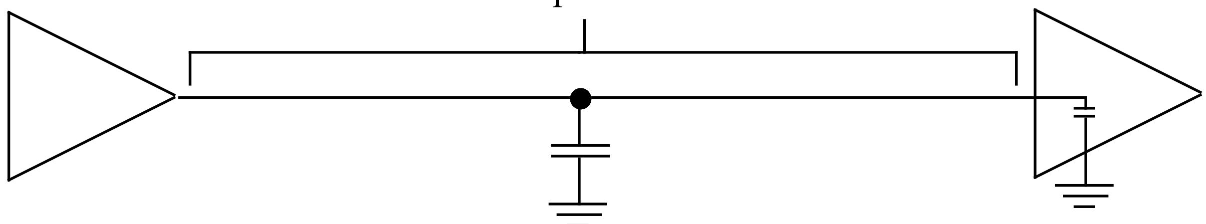
- SSI Logic and Cache Ram Propagation Delay Derated by 1ns

RIORDAN 2

Address Capture by IClk and DClk - This constraint guarantees that the transparent latches on the AdrLo bus close before the processor changes the address.



OUTPUT DERATING

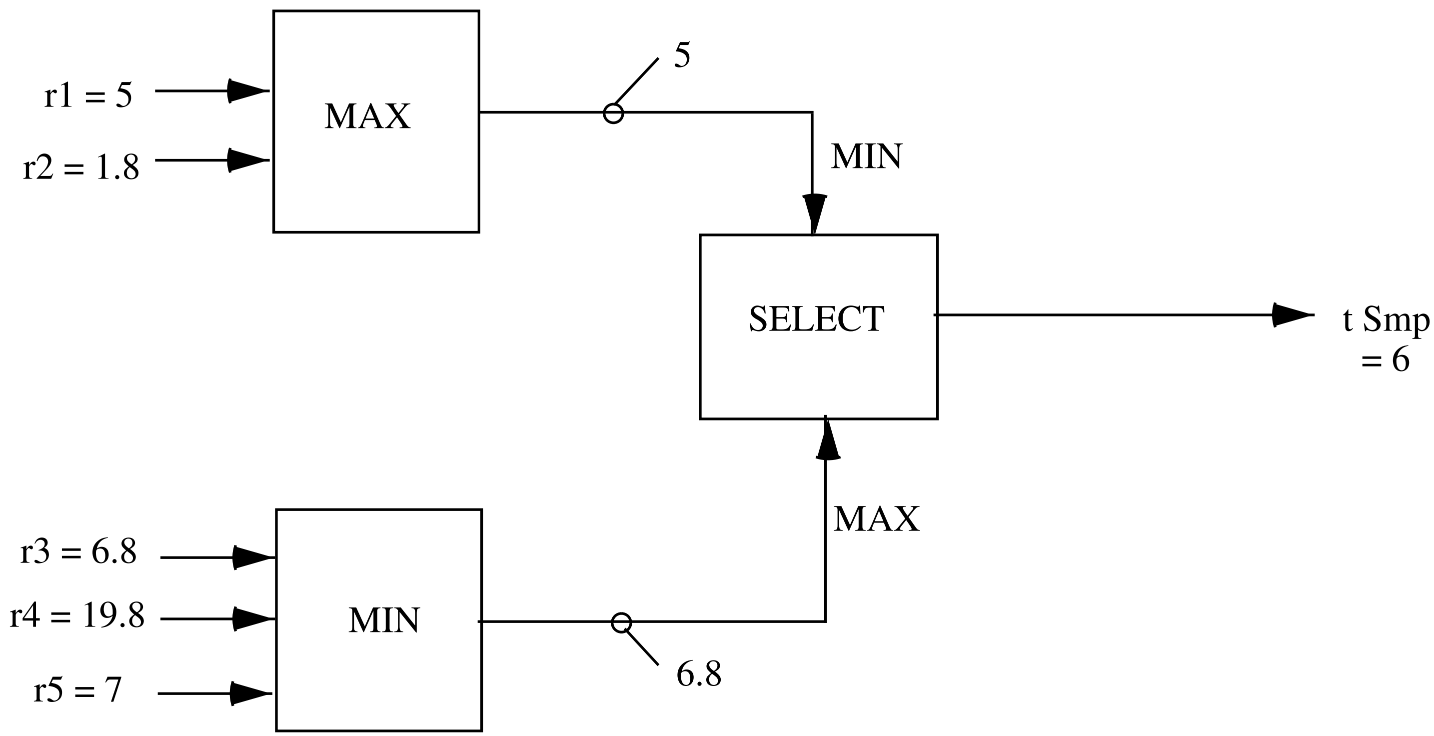


Derate 1ns/25pF beyond 25pF (Standard Load)

Parameter	ns
t374OeDe	0
tRamAaDe	1
tRamDoeDe	1
tDValDe	1
tIRdDe	1

Parameter	MIN	MAX
tClkDe	0.5	1
tAdrLoDe	0.5	1
tDRdDe	0.5	1
tRdDe	0.5	1
tSysDe	0.5	1
tWrDe	0.5	1
tXEnDe	0	1

CALCULATING t Smp

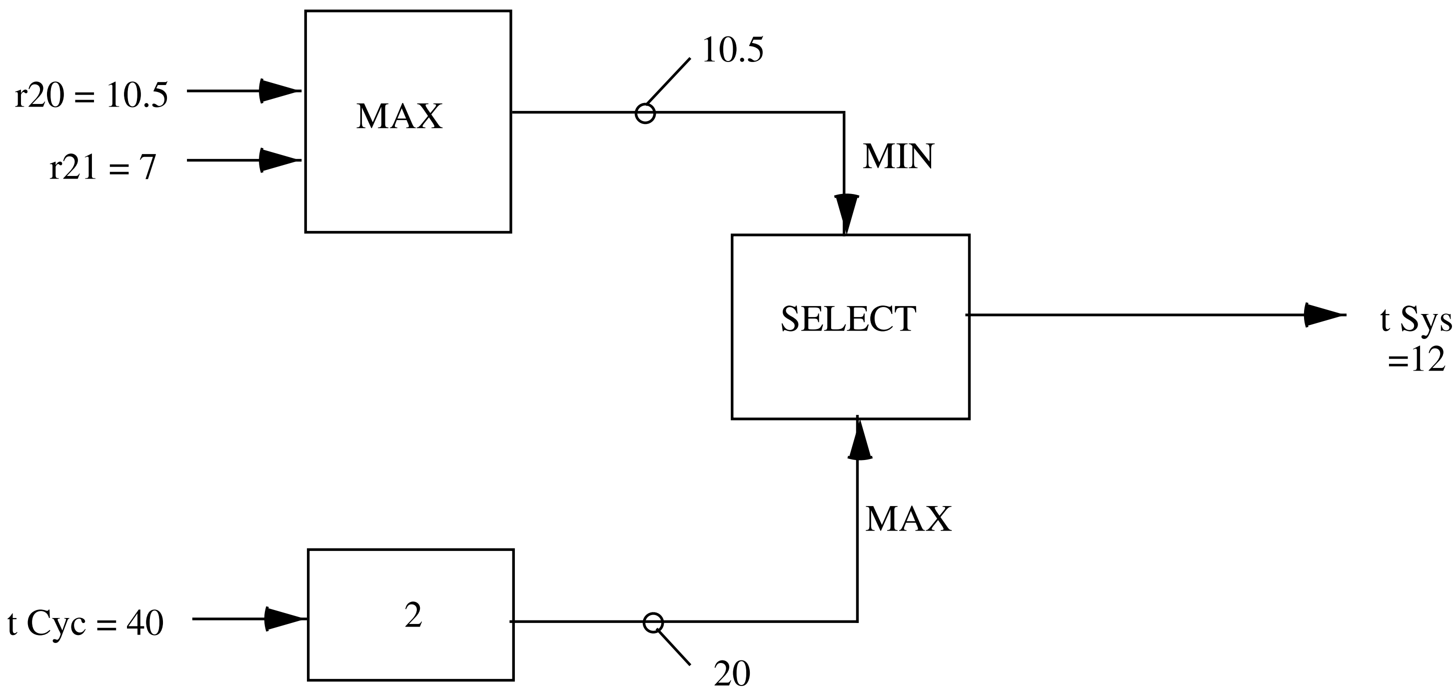


- Round Up the Min to the Next Even Number

CALCULATING t Sys

FORMULA	DESCRIPTION	FOR 25MHz ns
r20	Read-Write, ICache-Data Bus Contention	10.5
r21	Read-Write, Read Buffer-Data Bus Contention	7

CALCULATING t Sys

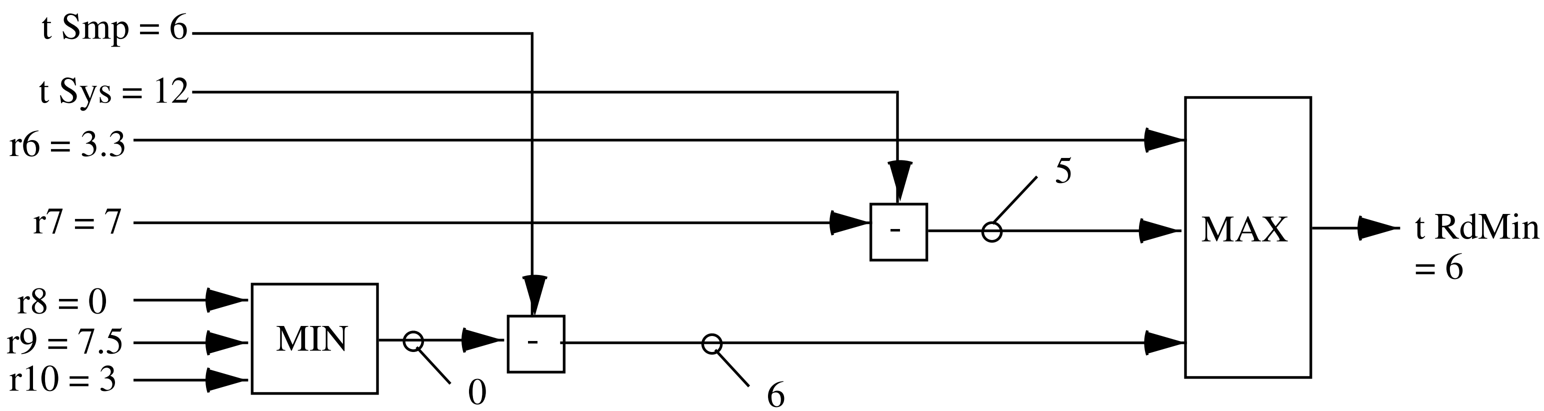


- Round Up the Min to the Next Even Number

CALCULATING t Rd Min

FORMULA	DESCRIPTION	FOR 25MHz ns
r6	Min Delay Thru Transparent Latches	t Rd 3.3
r7	Min Read Pulse Width	t Sys - t Rd 7
r8	Cache Enable to Sample	t Smp - t Rd 0
r9	Read Buffer Enable to Sample	t Smp - t Rd 7.5
r10	Read Buffer Data Setup to End of Write	t Smp - t Rd 3

CALCULATING t RdMin

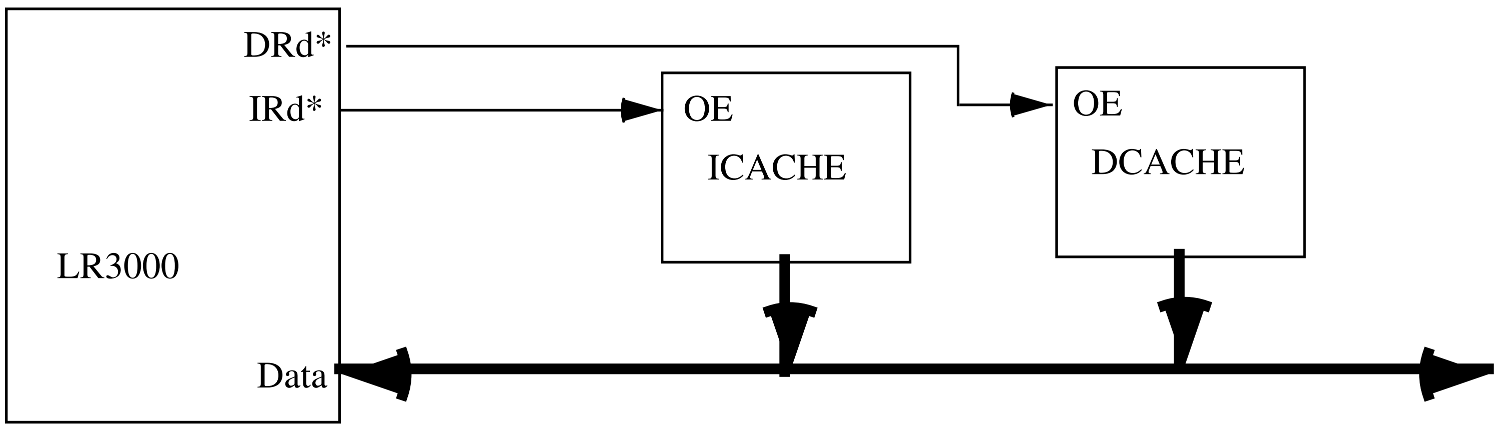


CALCULATING t Rd Max

FORMULA	DESCRIPTION	FOR 25MHz ns
r11	Data Hold from End of Write	t Smp - t Rd -2
r12	Address Hold from End of Write	t Smp - t Rd -2
r18	Read-Read, ICache-DCache Contention	t Sys - t Rd 1.6
r19	Read-Read, Read Buffer-DCache Contention	t Sys - t Rd 3

RIORDAN 18

Read - Read, ICache - DCache contention - This constraint guarantees no contention between the instruction cache and the data cache on a load.

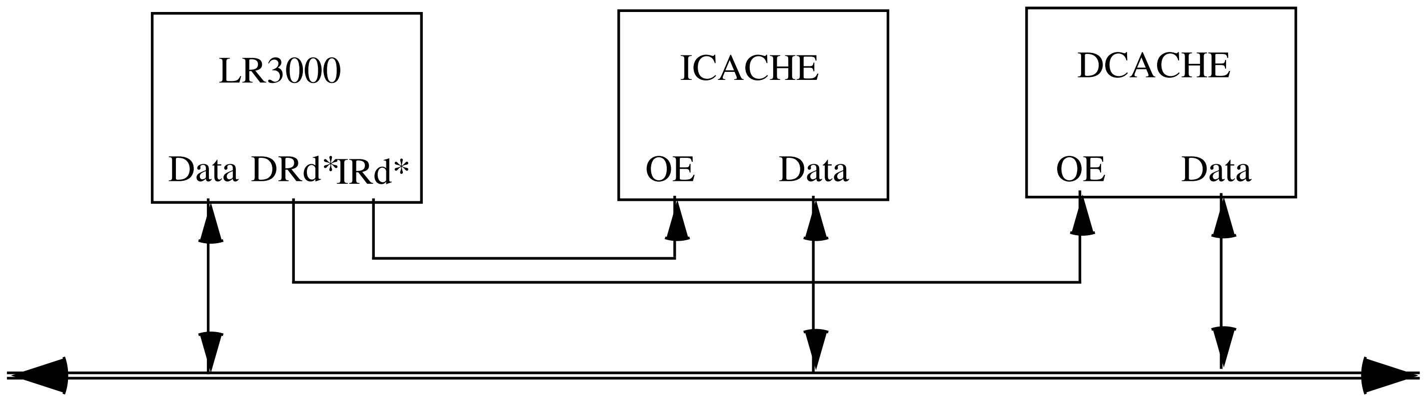


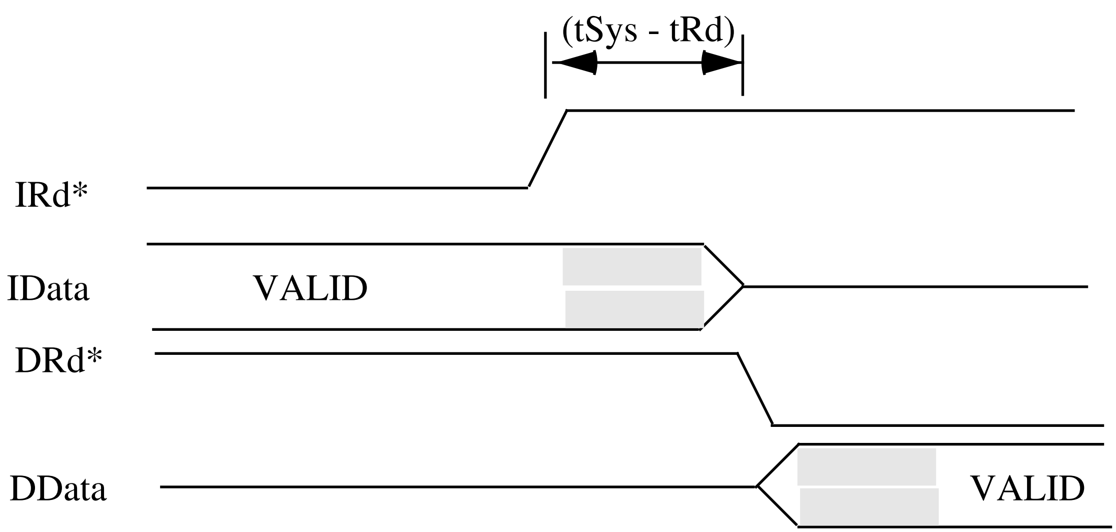
t Sys - t Rd   tIRdDe + tRamEDM - tDRdDeMax

= 1 + 1.6 - 1

= 1.6

WHAT IS t EDM ?
(ENABLE DISABLE MISMATCH)





- Outputs Must Not Overlap

<u>WHAT IS tEDM ?</u>

TOSHIBA TC 55417P/J - 20H

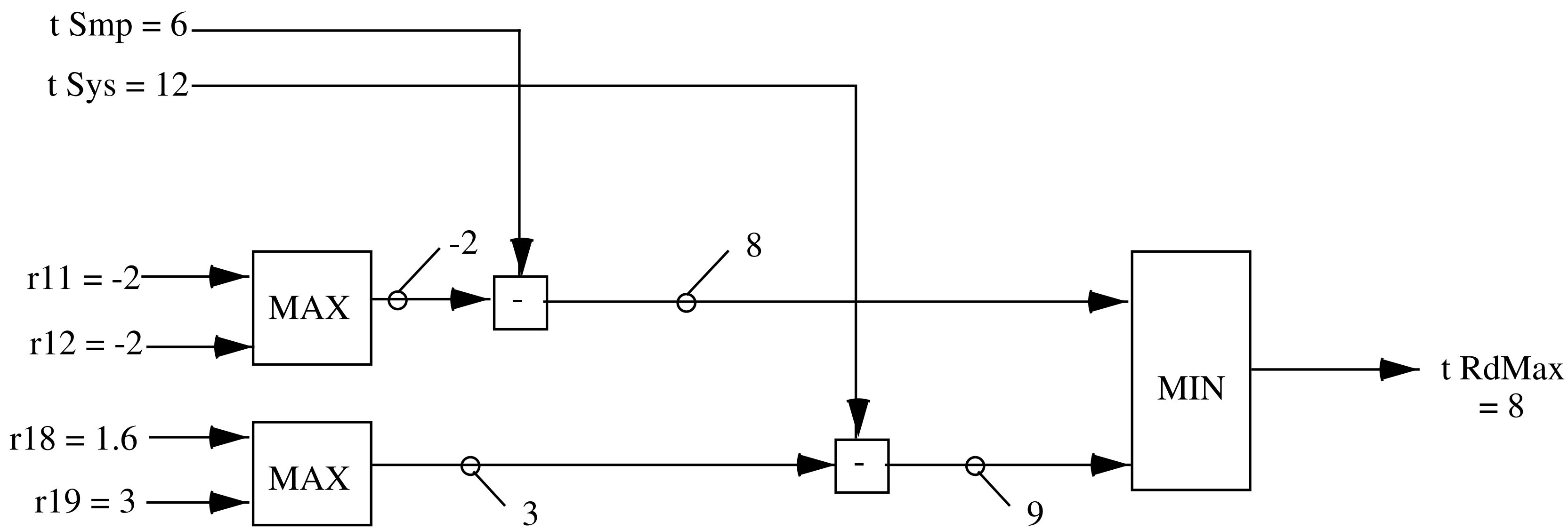
SYMBOL	PARAMETER	MIN	MAX	UNIT
tOEE	Output Enable to Output in Low-Z	0	-	ns
tODO	Output Enable to Output in High-Z	-	8	ns

0 °C 4.5V

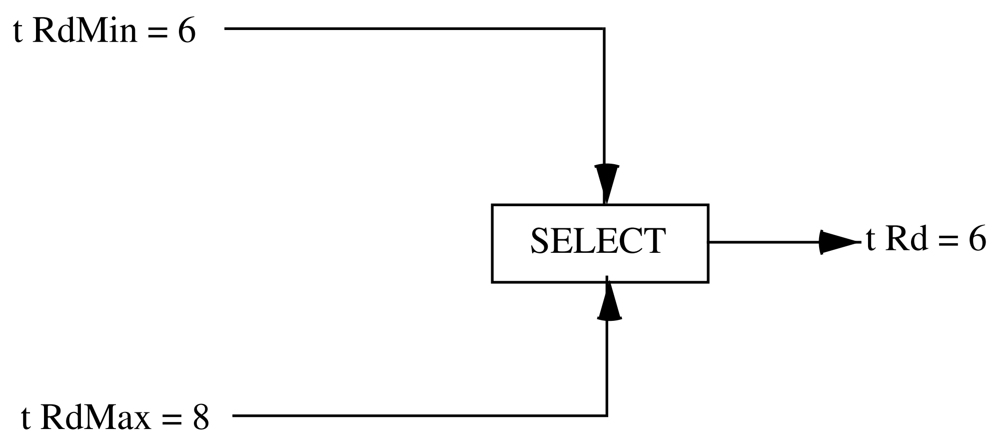
70 °C 5.5V

- There are no Specs for Same Test Conditions
- Estimated Value for tEDM is 1.6ns Allowing for Reasonable Variations in Manufacturing

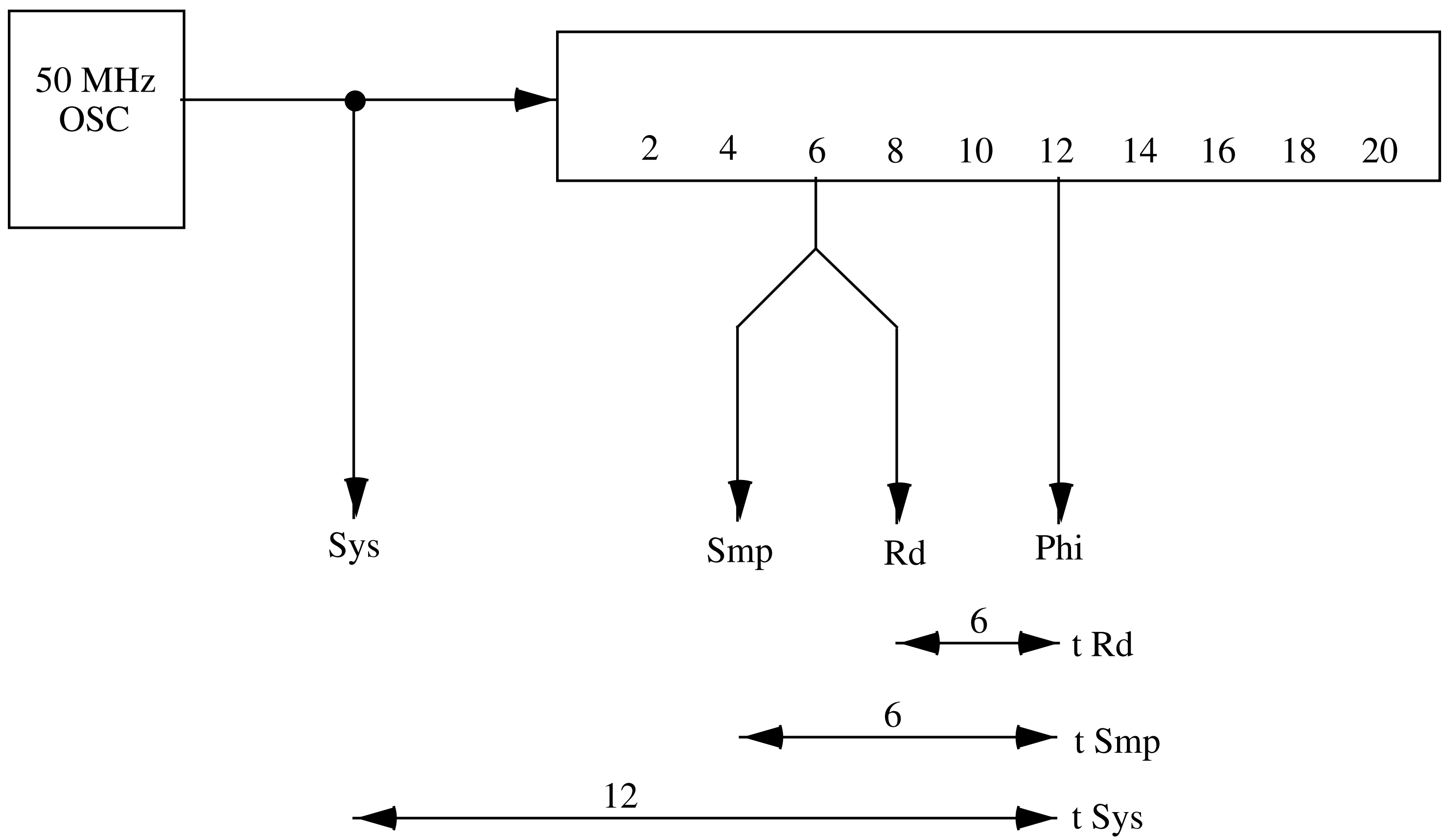
CALCULATING t RdMax



CALCULATING t Rd



SELECTING THE DELAY TAPS



THREE STEP RECIPE FOR 25MHz

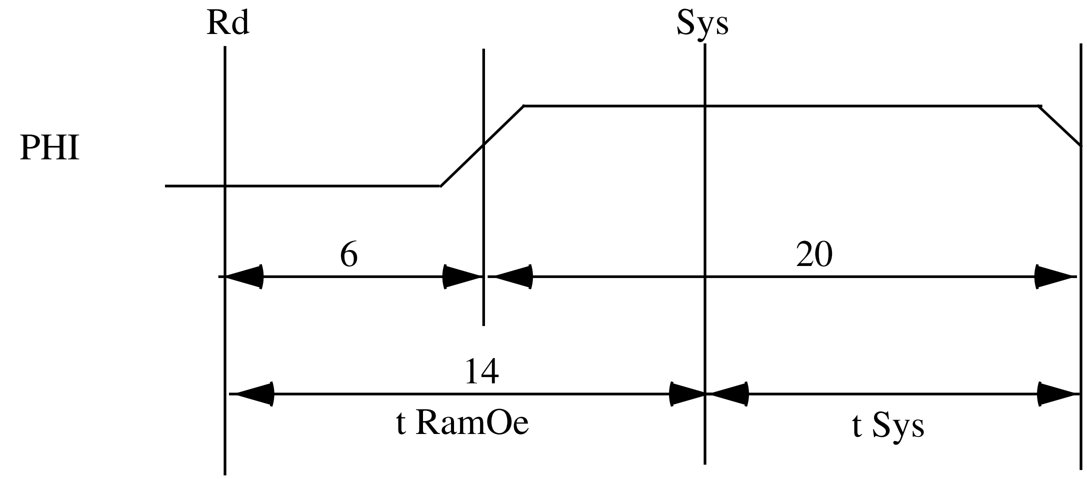
1. SET t Smp to 6ns
   - Guarantees Data Stable
   - Allows Parity, Tag Check, Load Align to Overlap Before PHI

2. SET t Rd to t Smp
   - UNLESS t RamHd > t RamLzMin

3. SET t Sys to t RamOe after Rd
   - Check t Sys > t RamHz + 2.5ns

THREE STEP RECIPE FOR 25MHz

3. SETTING t Sys to t RamOe after Rd



$$
\begin{aligned}
t\ RamOe &= tRamDoe + tRamDoeDe + Rounding \\
&= 12 + 1 + 1 \quad \text{(Round to next even value)} \\
&= 14 \\
t\ Sys &= ((t\ Cyc/2) + t\ Rd) - tRamOe \\
&= 20 + 6 - 14 \\
&= 12
\end{aligned}
$$

THREE STEP RECIPE FOR 25MHz

VERIFICATION

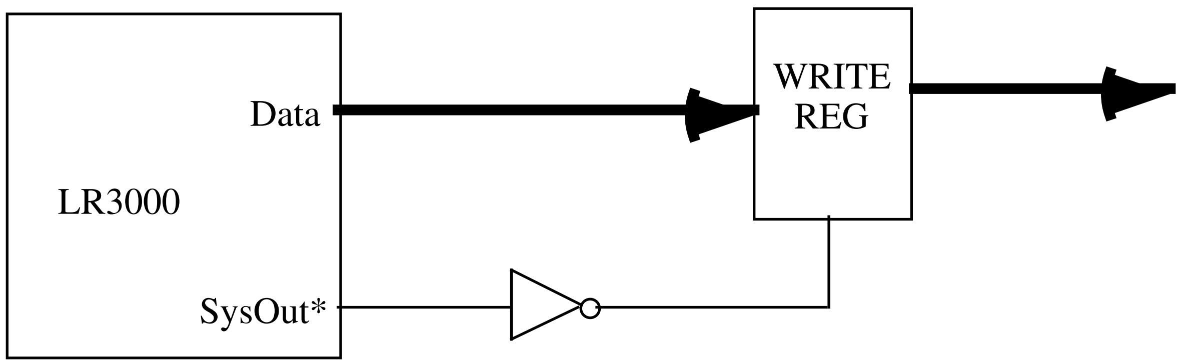
STEP 2: t RamHd MUST BE   t RamLzMin

0   2 $\boxed{\text{OK}}$

STEP 3: t Sys MUST BE > t RamHzMax + 2.5

12 > 8 + 2.5
12 > 10.5 $\boxed{\text{OK}}$

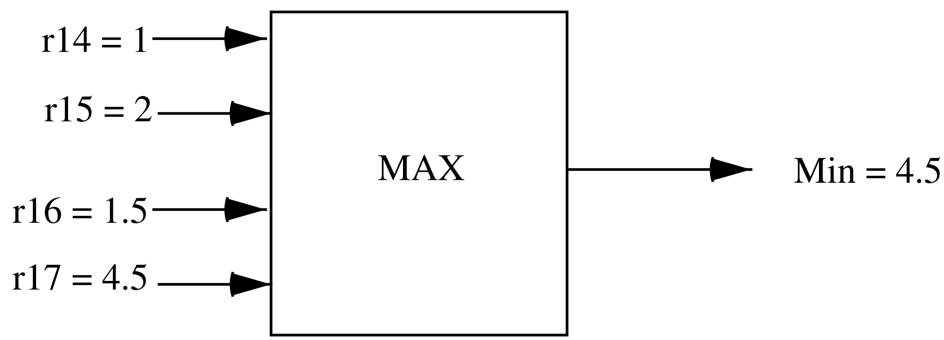
WRITE REGISTER TIMING



Setup Max = r25

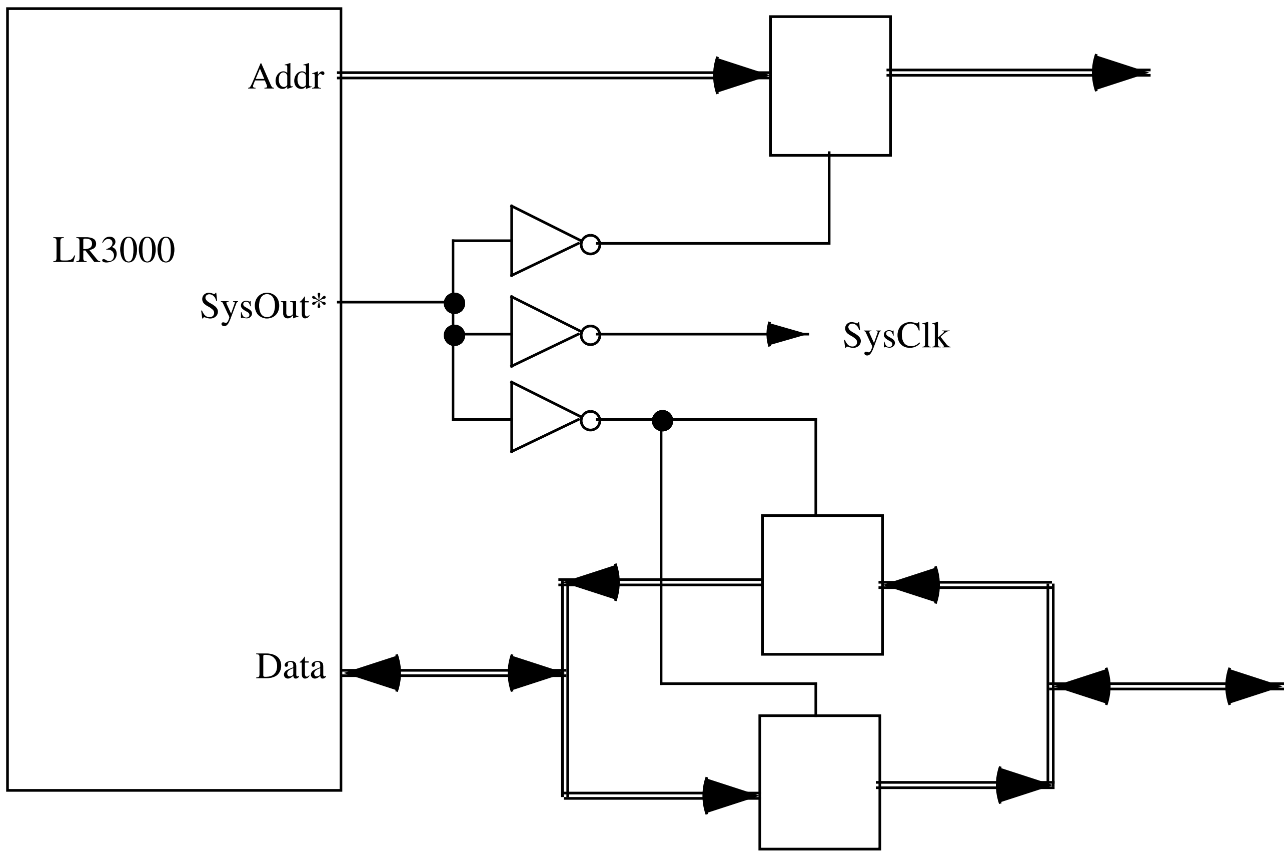
Hold Max = r26

CALCULATING DATA BUS HOLD TIME



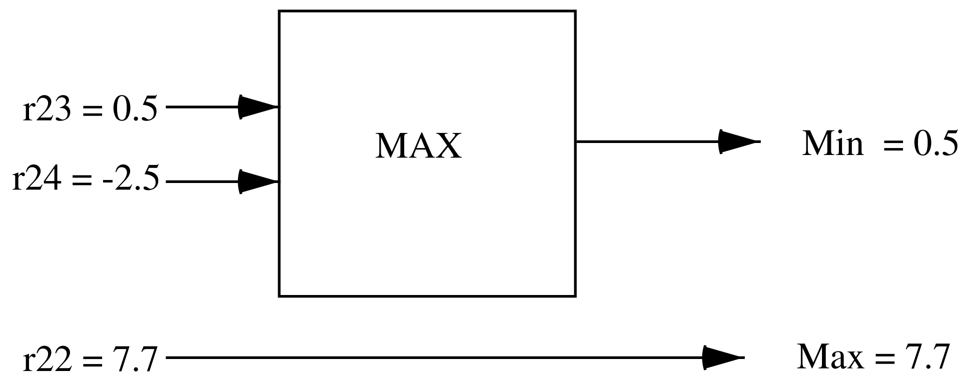
- Determines Minimum Time for Which Data Must Remain Valid on the Bus After it Stops Being Driven

CALCULATING SysClk DELAY



- Specifies Max & Min Delay for Inverter Used to Generate SysClk from SysOut*

CALCULATING SysClk DELAY



Part	Min	Max
FCT 1804A	1	4
FCT240A	1.5	4.8
74AS04	1	5