实践是最好的课堂

——龙芯处理器设计的启示

胡伟武

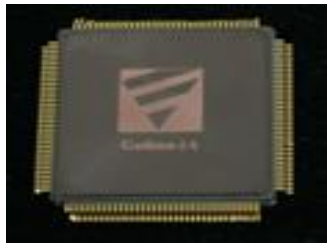
提　纲

- 龙芯处理器简介
- 知耻而后勇的性能提高过程
  - **Simulation-Silicon Correlation**（硅是检验设计的唯一标准）
  - **Balanced Design**（结构设计要统筹兼顾）
  - **Optimization**（结构设计要重点突出）
  - **Pico-Architecture Design**（面向工艺的结构设计）

龙芯处理器简介

持续改进的过程

- 龙芯1号：有了
  - 2001年5月正式启动龙芯CPU的研制
  - 2002年8月研制成功龙芯1号是我国第一个通用处理器芯片
- 龙芯2号：积累
  - 处理器的每年性能提高三倍；
  - 龙芯2E/2F主频1GHz，在64位单处理器设计方面达到世界先进水平
  - 龙芯2F批量生产，几十个应用
- 龙芯3号：跨越
  - 四核龙芯3A流片成功并量产、8核龙芯3B流片成功
  - 形成自己的特色和竞争优势
  - HotChips, IEEE Micro, ISCA、HPCA、ISSCC等国际著名刊物和会议发表龙芯3号结构



X3

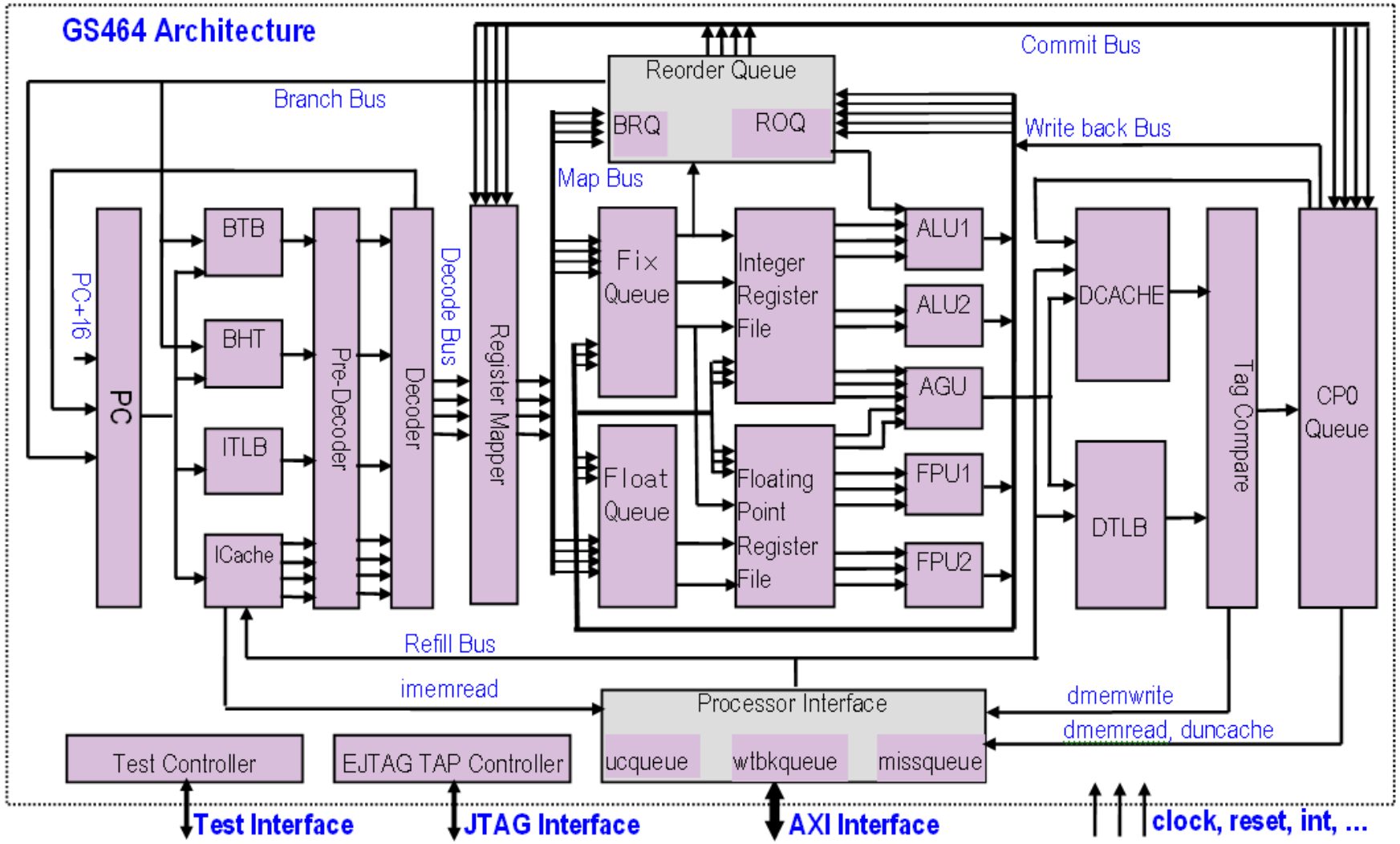
X3

X3

GS464处理器核特点

- MIPS64兼容，增加SIMD型多媒体指令以及X86虚拟机指令
- 四发射超标量结构，两个定点、两个浮点、一个访存部件
- 每个浮点部件可扩展成256位SIMD部件
- 访存部件支持128位存储访问，虚地址和物理地址各为48位
- 支持寄存器重命名、动态调度、转移预测等乱序执行技术
- 64项全相联TLB，独立的16项指令TLB，可变页大小
- 一级指令Cache和数据Cache大小各为64KB，4路组相联
- 支持Non-blocking访问及Load-Speculation等访存优化技术
- 支持Cache一致性协议，可用于片内多核处理器
- 指令Cache实现奇偶校验，数据Cache实现ECC校验
- 支持标准的EJTAG调试标准，方便软硬件调试
- 标准的128位AXI接口

龙芯2号处理器核（GS464）



龙芯处理器结构设计心得

- 短短5年走过了处理器结构设计近二、三十年的发展历程，难得有这样的经历
- 结构设计可以采用跨越的方法（如龙芯1号采用动态调度、龙芯2号采用四发射乱序执行），但认识的提高，经验的增长是无法跨越的
- “小步快跑”的技术路线加速了认识提高和经验增长的过程
- 寓乐于苦，每个芯片流片成功只能有1-2天的短暂欢乐，但是够了

Table 1. SPEC_int2000 and SPEC_fp2000 of Godson Processors				
SPEC Programs	Godson-1(200MHz)	Godson-2B(250MHz)	Godson-2C (450MHz)	Godson-2E (1GHz)
164.gzip	18.3	39	104	347
175.vpr	20.8	57	169	512
176.gcc	13.3	39	103	497
181.mcf	49.4	64	156	586
186.crafty	7.6	36	234	598
197.parser	24.1	49	136	382
252.eon	11.3	57	287	690
253.perlbmk	17.2	53	172	508
254.gap	20.2	67	135	458
255.vortex	12.4	43	180	722
256.bzip2	29.1	67	155	411
300.twolf	25.2	68	158	465
SPEC_INT2000	18.5	52	159	503
168.wupwise	35.6	88	145	672
171.swim	37.7	55	116	469
172.mgrid	14.7	40	69	311
173.applu	24.3	45	77	382
177.mesa	22.7	100	215	634
178.galgel	-	-	-	704
179.art	55.3	66	125	624
183.equake	35.5	58	100	624
187.facerec	-		-	632
188.ammp	12.3	62	147	509
189.lucas	-	-	-	506
191.fma3d	-	-	-	395
200.sixtrack	16.6	46	92	319
301.apsi	20.4	46	115	493
SPEC_FP2000	24.8	58	114	503

龙芯3A3000的SPEC CPU2000性能

- 编译优化后peak分值（gcc+lcc）

Benchmark	Reference Time	Base Runtime	Base Ratio
164.gzip	1400	214	653
175.vpr	1400	146	957
176.gcc	1100	74.1	1485
181.mcf	1800	68.8	2618
186.crafty	1000	68.9	1452
197.parser	1800	179	1007
252.eon	1300	123	1055
253.perlbmk	1800	196	918
254.gap	1100	104	1057
255.vortex	1900	80.4	2362
256.bzip2	1500	139	1077
300.twolf	3000	243	1234
SPECint_base2000			1225

Benchmark	Reference Time	Base Runtime	Base Ratio
168.wupwise	1600	50.2	3188
171.swim	3100	122	2546
172.mgrid	1800	118	1526
173.applu	2100	79.1	2653
177.mesa	1400	89.5	1565
178.galgel	2900	49.9	5810
179.art	2600	15.8	16451
183.equake	1300	41.8	3109
187.facerec	1900	108	1753
188.ammp	2200	180	1219
189.lucas	2000	88.0	2273
191.fma3d	2100	137	1530
200.sixtrack	1100	151	730
301.apsi	2600	130	2006
SPECfp_base2000			2360

3A3000的SPEC CPU2006性能

- 测试运算性能和访存带宽，均使用gcc编译器的最高优化选项
  - SPEC INT2006测试定点运算性能，SPEC FP2006测试浮点运算性能
  - STREAM测试访存带宽

	龙芯3A3000	VIA-C4600	AMD K10
主频	1.5GHz	2.0GHz	1.5GHz
核数	4	4	4
SPEC INT2006 ratio	11.1	10.8	11.3
SPEC INT2006 rate	36.2	27.5	36.6
SPEC FP2006 ratio	10.1	9.8	11.3
SPEC FP2006 rate	32.9	23.2	34.0
Stream copy单核带宽	8.8	4.5	4.5
Stream copy四核带宽	13.2	3.3	6.0

硅是检验设计的唯一标准

硅是检验结构设计的唯一标准

- 模拟器是处理器结构设计的重要平台
  - 龙芯1号和龙芯2号采用“可执行”的结构设计的理念，以模拟器作为结构设计的文挡
- FPGA验证是龙芯流片前的支柱性验证平台
  - 可以在真实的主板上运行，因而更加准确
- 模拟器和FPGA在性能分析方面的欺骗性
  - 影响性能的结构参数相当复杂，模拟和仿真一般集中在几个“重要”参数
  - 设计人员的经验不足导致参数的设置不准确
  - 设计人员的良好愿望导致倾向性，忽略对自己不利的因素，挖掘对自己有利的因素

故事1：模拟器和硅的校准

- 龙芯1号FPGA性能可比50MHz的Intel 486
  - 预期200MHz的龙芯1号性能可比200MHz的P2
- 500MHz龙芯2C的性能预期为1GHz的P3
  - 500MHz的MIPS R10000和Alpha 21264能够达到
- 实际都比预期的性能低1倍
  - FPGA的访存延迟过于乐观
  - 龙芯2C的项目没有验收
- 龙芯2E比龙芯2C主频提高1倍，性能提高2倍

测试程序	运行时间（秒）			相对时间		
	Godson	486	IDT	Godson	486	IDT
浮点矩阵乘法	1.99	2.68	0.37	1.00	0.99	0.18
FFT	52.89	56.02	7.97	1.00	1.06	0.15
SOR	5.59	5.60	0.91	1.00	1.00	0.16
计算π	58.03	164.71	11.57	1.00	2.84	0.20
Whetstone	9.11	7.15	1.07	1.00	0.78	0.12
定点矩阵乘法	50.61	28.08	16.61	1.00	0.55	0.33
164.gzip	41.21	24.17	5.49	1.00	0.59	0.13

FPGA和硅片的参数校准

- 在龙芯2C中，对FPGA和真实芯片性能进行了认真的校准
  - 即便如此，部分程序仍有10%的差距
- 龙芯2E在FPGA阶段对SPEC CPU2000分值进行评估，使用了比较保守的延迟参数后，性能仍然高估10%-20%
  - 忽略了访存冲突
  - 良好的愿望

Table 3 SPEC CPU2000 Correlation between Silicon and FPGA

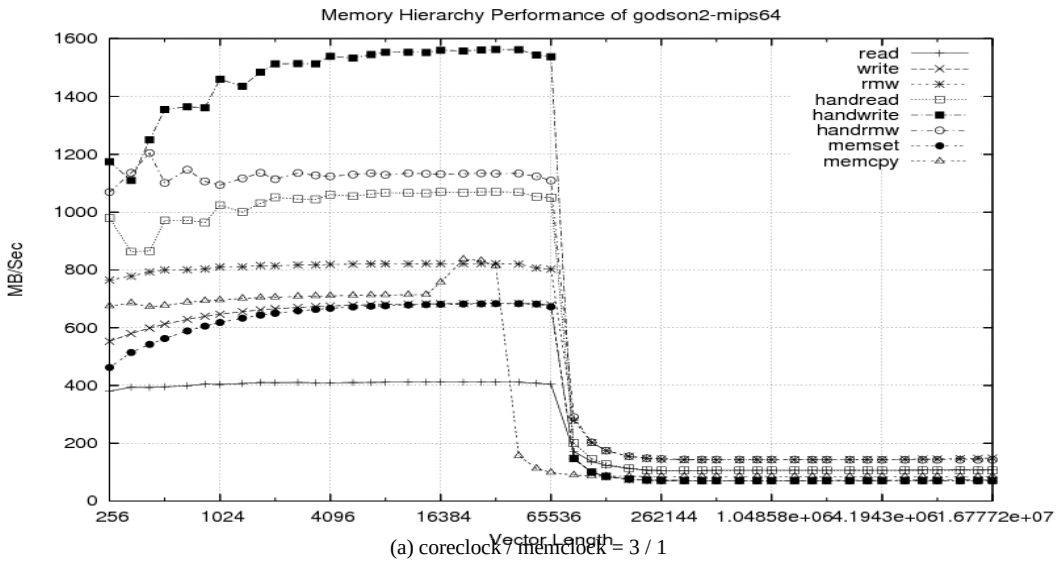
	2C (test data set)		2E (train data set)	
	Real(s)	FPGA(s)	Real(s)	FPGA(s)
164.gzip run time(s)	23.9	23.4	52.5	53
175.vpr run time(s)	20.0	19.9	30.5	30.8
176.gcc run time(s)	25.1	24.6	5.3	5.4
181.mcf run time(s)	3.91	3.79	48.1	41.6
186.crafty run time(s)	73.1	69.3	25.3	26.9
197.parser run time(s)	37.3	35.9	12.5	12.1
253.perl run time(s)	19.3	19.0	85.6	93.4
254.gap run time(s)	11.4	11.3	11.0	10.2
255.vortex run time(s)	134	144.6	18.6	19.6
256.bzip2 run time(s)	58.9	60.9	63.4	68.5
300.twolf run time(s)	1.92	2.10	-	20.5
168.wupwise run time(s)	66.1	63.0	63.2	53.1
171.swim run time(s)	11	10.6	26.9	15.2
172.mgrid run time(s)	213	229	37.4	28.9
173.applu run time(s)	4.28	4.06	26.8	19.0
177.mesa run time(s)	15.7	16.0	67.4	76.2
179.art run time(s)	96.7	91.3	27.4	13.6
183.equake run time(s)	12.3	12.1	59.4	42.9
188.ammp run time(s)	130	127	70.7	80.6
200.sixtrack run time(s)	96.5	94.3	133.9	161.6
301.apsi run time(s)	82.4	78.9	19.5	16.7

Table 2 Memory Access Correlation between Godson-2C and FPGA

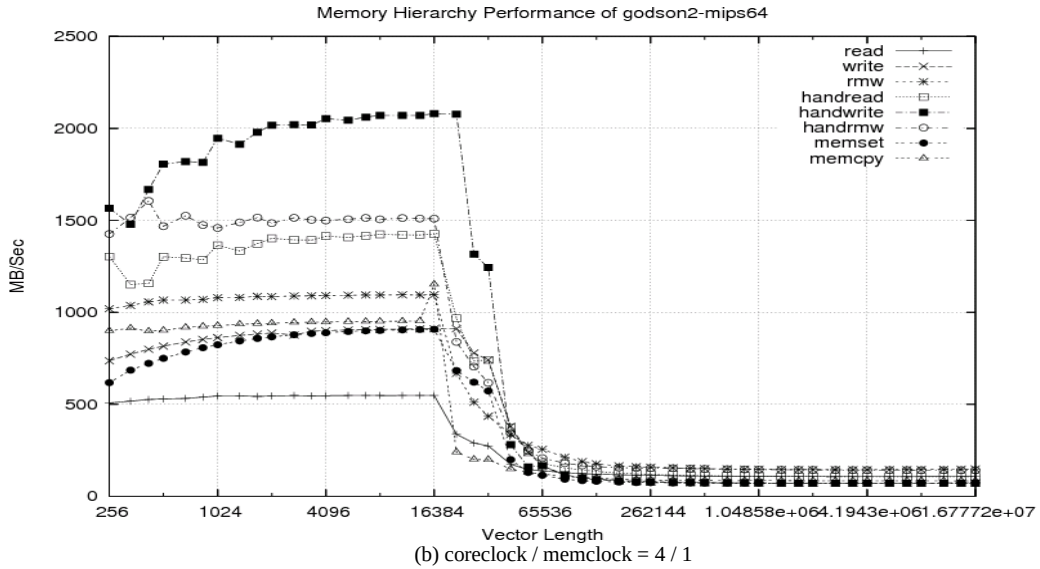
	Read (cycles)	Write (cycles)	Copy (MB/s)	Scale (MB/s)	Add (MB/s)	Triad (MB/s)
Godson-2C	26.00	37.88	71.88	70.00	75.12	76.09
FPGA	25.97	37.98	71.70	70.30	76.20	77.20

故事2：细节决定成败

- 龙芯2C的四路Cache随机替换计数器问题
  - 2倍频和4倍频的时候四路变成二路



(a) coreclock / memclock = 3 / 1

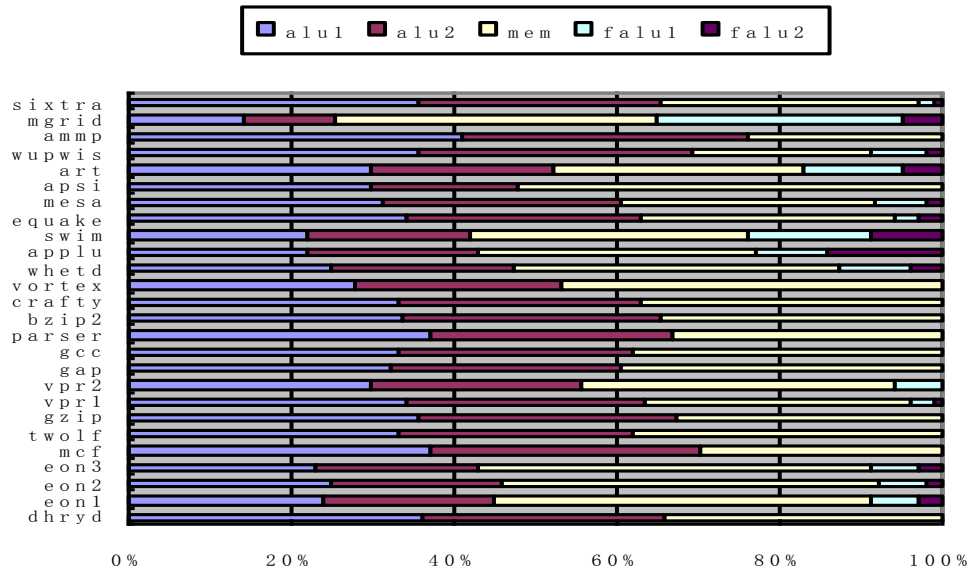


(b) coreclock / memclock = 4 / 1

Figure 2. Bandwidth of Godson-2C machine with coreclock/memclock of 3/1 and 4/1

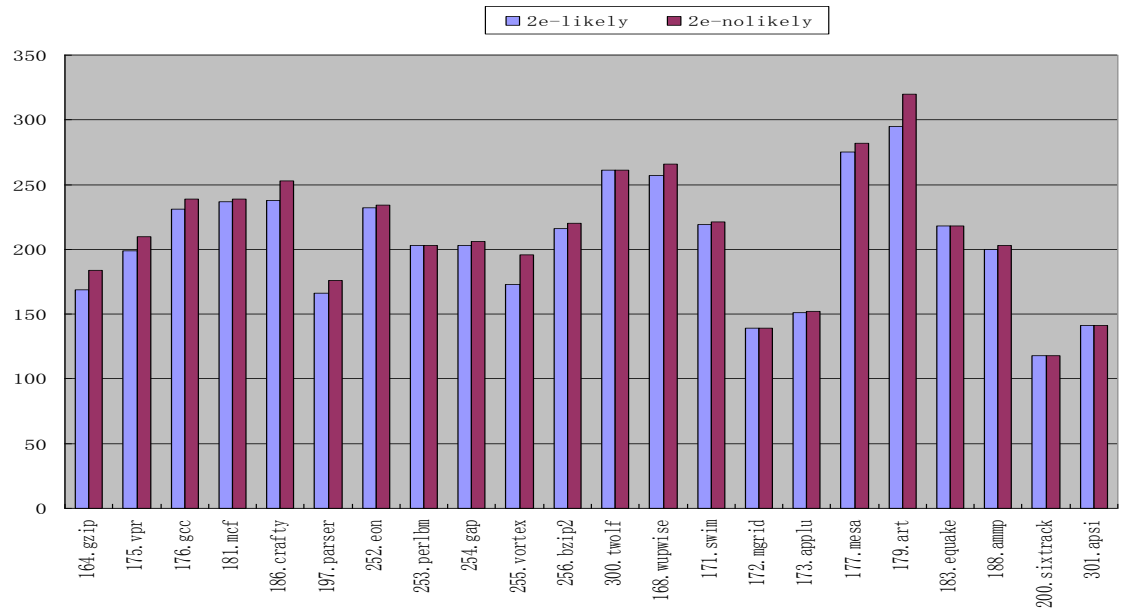
故事3：正确对待前人的做法和结果

- 功能部件的调度
  - **ALU1和ALU2公共指令：加减、逻辑运算**
  - **前人的结果移位指令使用不多，但gcc使用移位指令实现简单乘法**
  - **公共指令的调度：R10000根据目标寄存器号调度，导致不平衡**
  - **性能提高5%以上**



故事4：编译器和硬件的磨合

- 转移指令的编译
  - **Branch指令和Branch likely指令**
  - **硬件不猜Branch likely指令**
  - **编译器“滥用”branch likely指令**
  - **禁用branch likely指令后性能显著提高**

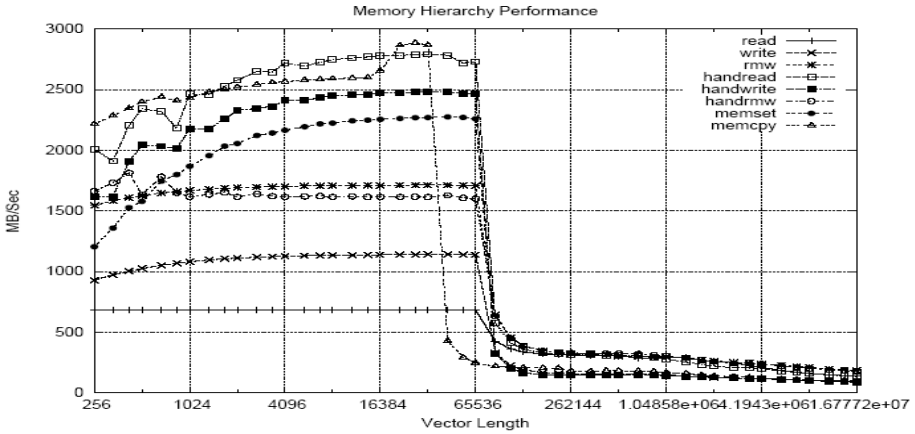


结构设计要统筹兼顾

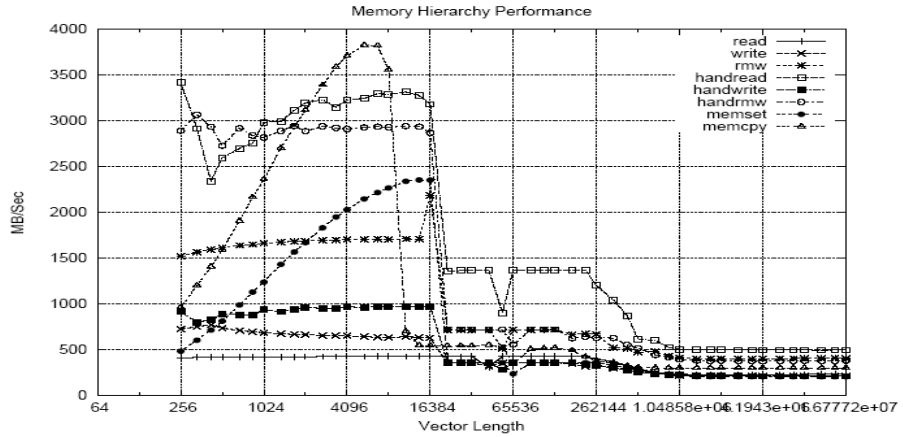
设计要统筹兼顾

- 一般的结构研究专注于考虑结构中某些重要因素的改善，但在一个通用的设计中，影响性能的因素非常复杂
- 一些次要因素往往成为整个设计的瓶颈，或当重要的瓶颈问题解决后，原来不是瓶颈的次要的因素成为瓶颈
  - 例如，龙芯1号的瓶颈在片内的L1和转移猜测，而龙芯2C在解决这些问题后，访存带宽成为系统瓶颈
- 设计者经验的缺乏也往往会忽视一些瓶颈问题，等到发现时，已经晚了
  - 象龙芯这样具有持续改善的机会的设计机会不多

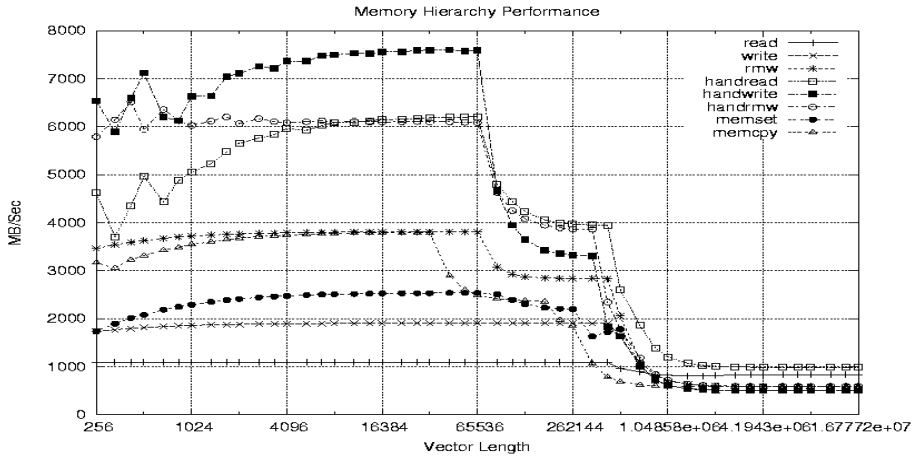
故事5：访存带宽瓶颈



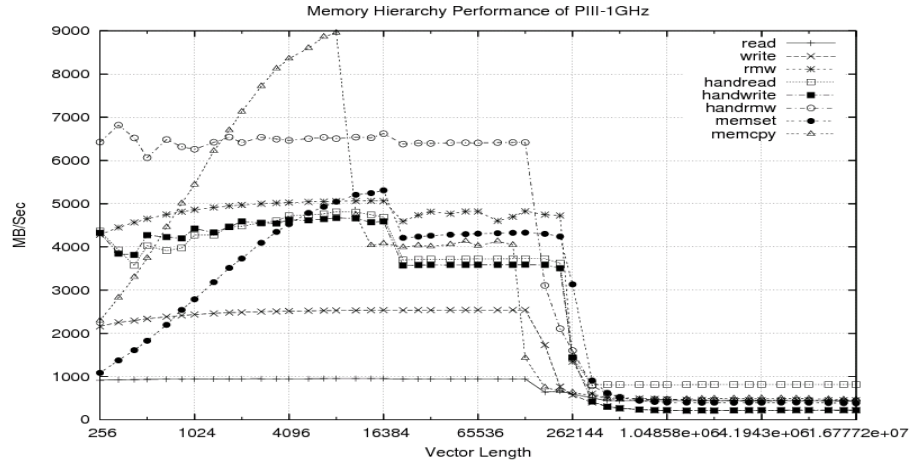
(a) 450MHz Godson-2C bandwidth



(b) 450MHz Pentium III bandwidth



(c) 1GHz Godson-2E bandwidth



(d) 1GHz Pentium III bandwidth

Figure 4. Memory bandwidth of Godson-2C, Godson-2E, and Pentium III

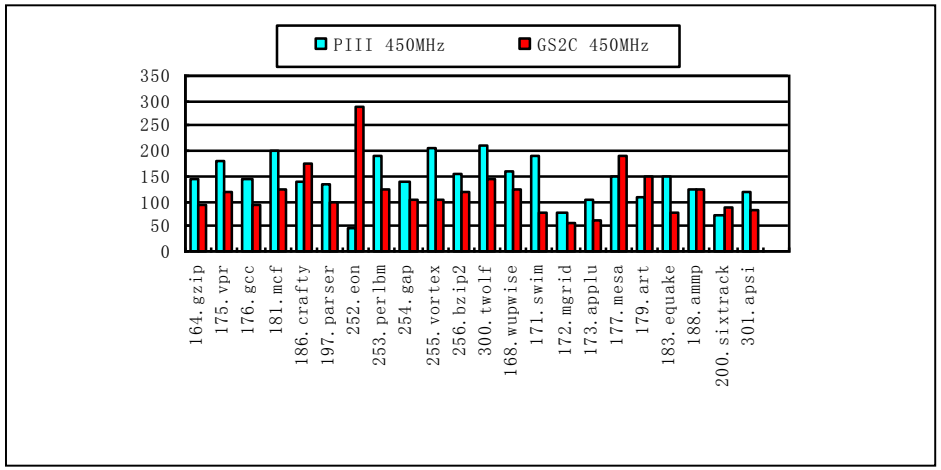
3A3000的访存带宽

- 测试运算性能和访存带宽，均使用gcc编译器的最高优化选项
  - SPEC INT2006测试定点运算性能，SPEC FP2006测试浮点运算性能
  - STREAM测试访存带宽

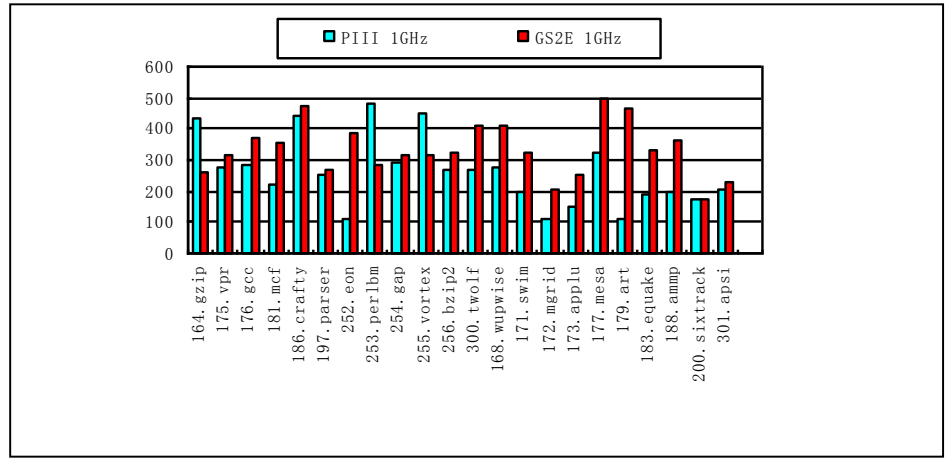
	龙芯3A3000	VIA-C4600	AMD K10
主频	1.5GHz	2.0GHz	1.5GHz
核数	4	4	4
SPEC INT2006 ratio	11.1	10.8	11.3
SPEC INT2006 rate	36.2	27.5	36.6
SPEC FP2006 ratio	10.1	9.8	11.3
SPEC FP2006 rate	32.9	23.2	34.0
Stream copy单核带宽（GB/s）	8.8	4.5	4.5
Stream copy四核带宽（GB/s）	13.2	3.3	6.0

龙芯2C/2E vs. Pentium III

- **龙芯2E的主频是龙芯2C的2倍，性能是龙芯2C的三倍，主要是由于L2和内存控制器在片内**



(a) Performance of Godson-2C and Pentium III

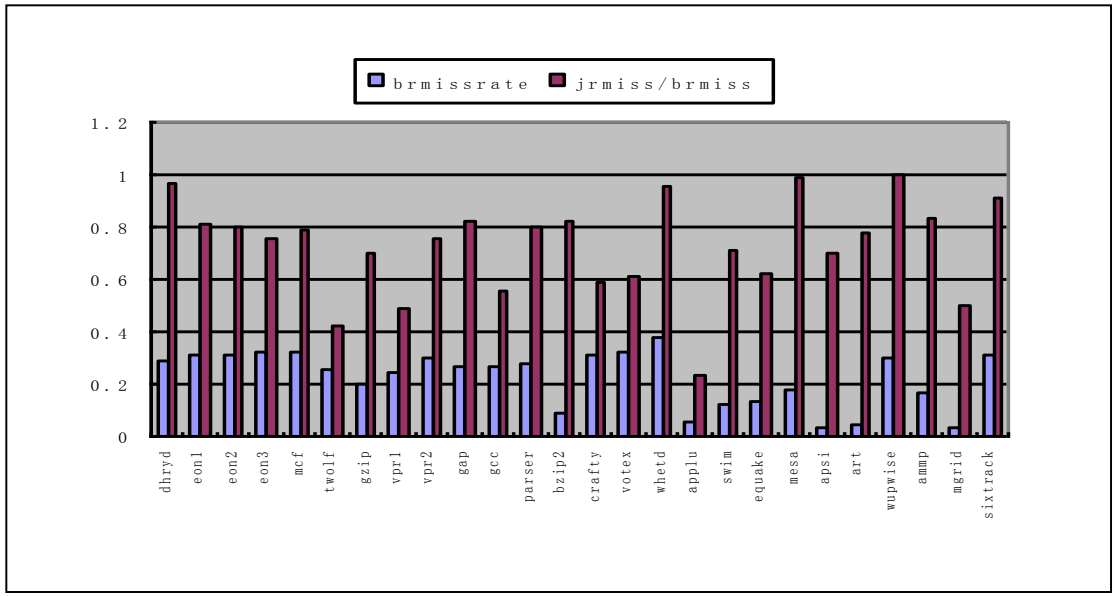


(b) Performance of Godson-2E and Pentium III

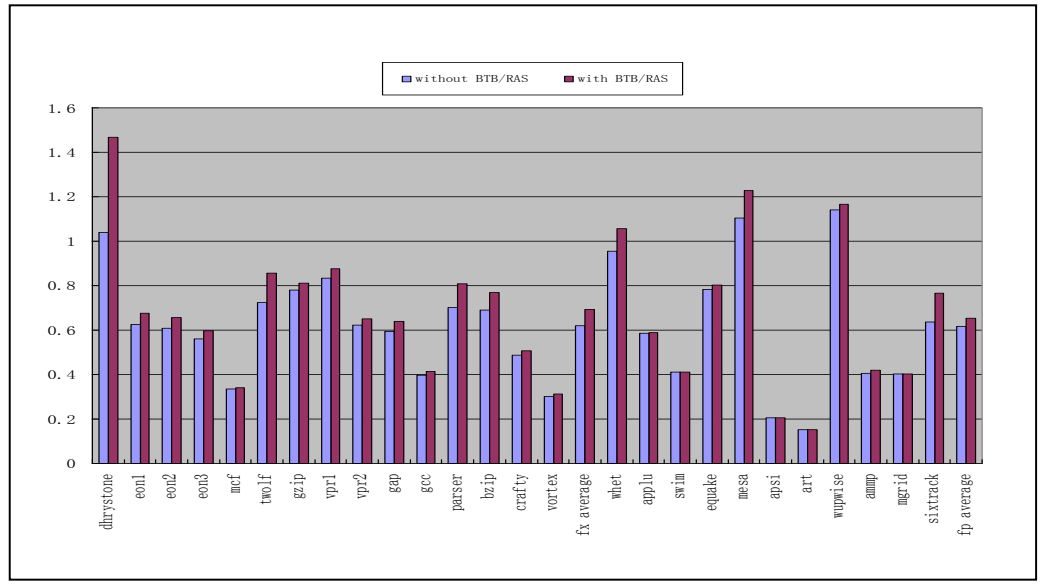
Figure 5. Performance Comparison of Godson-2C, Godson-2E and Pentium III (with GCC -O2 compiler)

故事6：间接转移猜测

- 转移猜测对处理器的性能至关重要
- 条件转移指令占90%，间接转移指令站10%
- 龙芯2B实现了2048项的PHT表用于条件转移猜测
- 龙芯2C/2E增加了16项BTB和4项RAS
  - 定点性能提高11.7%，浮点性能提高6.2%



(a) Branch misprediction rate



(b) IPC without and with BTB and RAS

Figure 6. Performance improvement caused by prediction of JR instructions

故事7: TLB例外开销

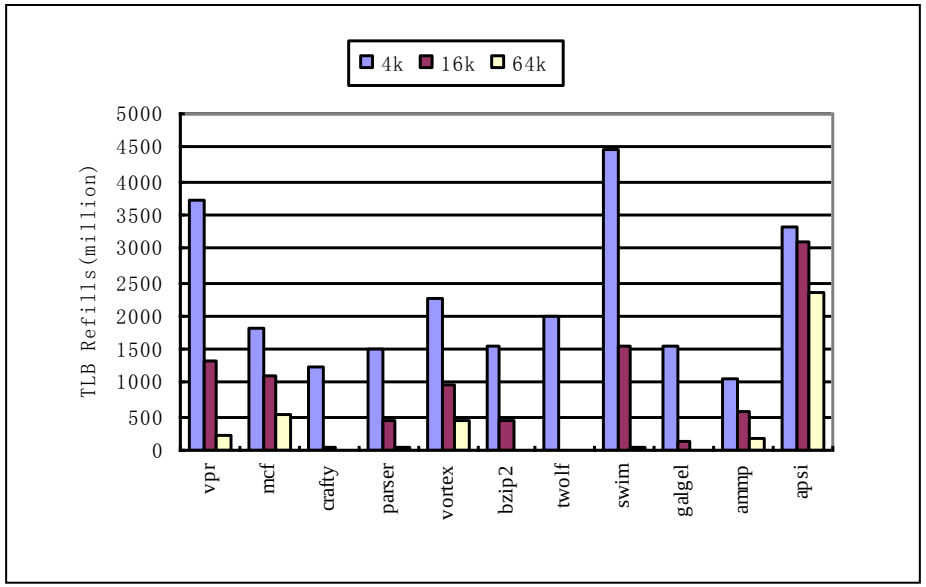
- 应用程序运行时，OS时间很少，主要是用户程序在“干活”
- 有些程序OS时间占30%以上，主要由于缺页例外处理造成
- TLB失效的概率远远小于Cache失效的概率，但由于失效开销大，仍成为系统瓶颈

Table 4. Operating System Overhead for part of SPEC CPU2000 programs

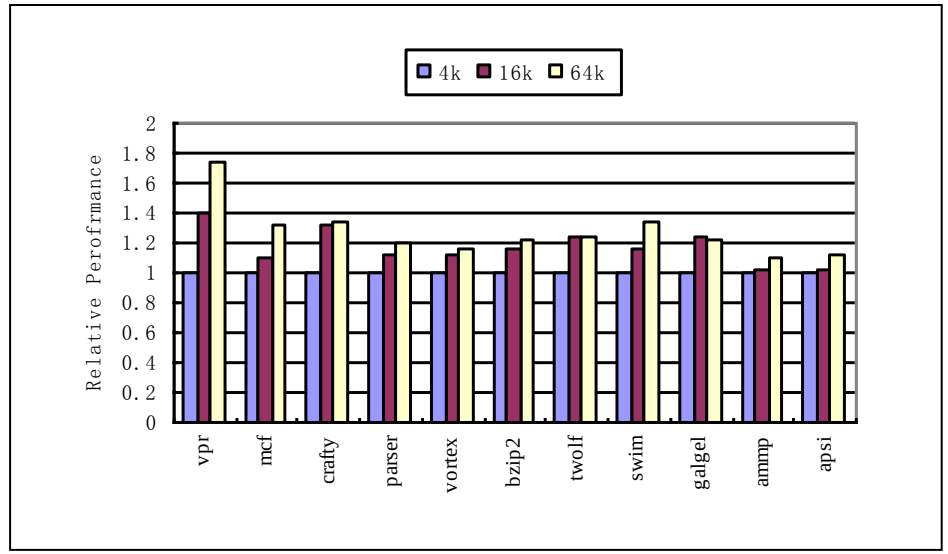
	Total cycles	OS cycles	OS percentage	TLB misses
Eon	941,904,123	8,478,817	0.90%	2,147
Parser	1,071,248,381	51,328,234	4.79%	98,976
Equake	1,071,008,004	68,737,907	6.42%	107,727
Gcc	1,072,539,177	73,025,370	6.81%	284,122
Mcf	681,765,369	63,143,562	9.26%	557,814
Crafty	1,072,161,097	125,172,455	11.67%	1,061,420
Mesa	1,071,869,442	196,088,410	19.29%	1,305,843
Vortex	1,072,127,164	214,648,139	20.02%	1,412,038
Bzip	1,071,228,969	326,063,247	30.44%	2,526,992

增加页大小后性能显著提高

- **TLB失效明显减少，16KB页时128页有2MB**
- **龙芯3号页大小动态可变**



(a) TLB refills of 4KB, 16KB, and 64KB page size



(b) Performance of 4KB, 16KB, and 64KB page size

Figure 7. TLB refills and relative performance of 4KB, 16KB, and 64KB page size

结构设计要重点突出

设计要重点突出

- 系统性能基本平衡后，要把有限的资源投入到优化效果最明显的地方：好钢用在刀刃上
- 优化在程序运行中最经常发生的事件
- 龙芯2号设计过程中进行了大大小小几百处优化
- 性能是一点点“抠”出来的

故事8: 降低Load-to-Use延迟

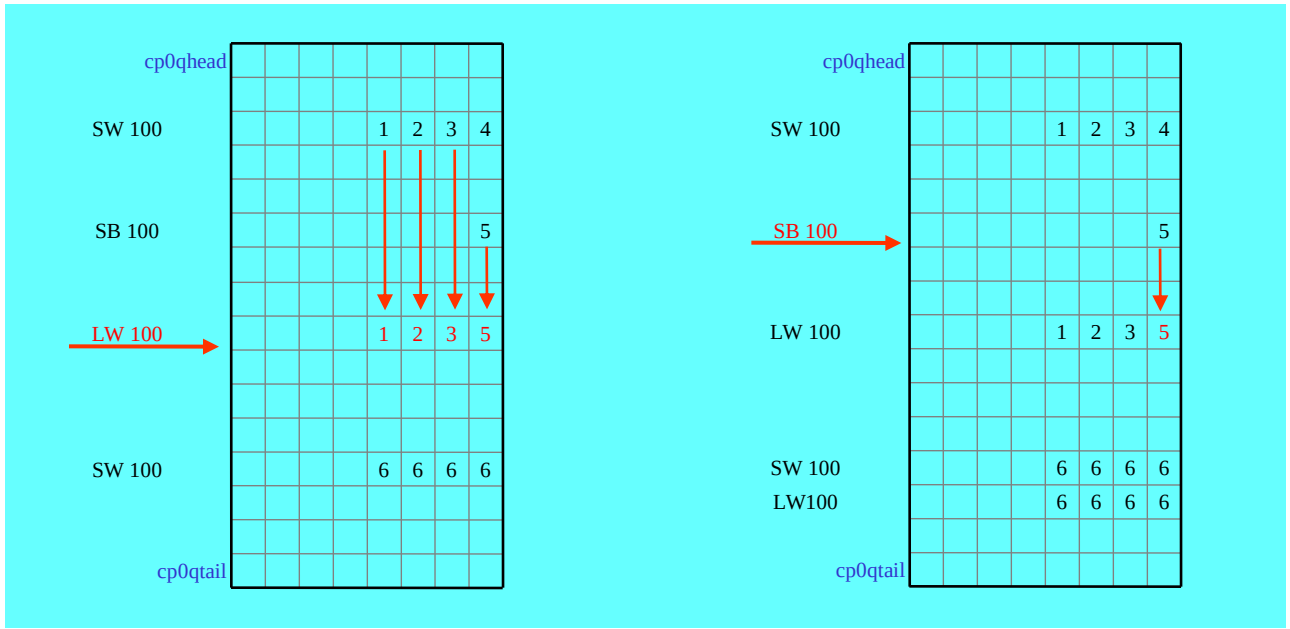
- **Load-to-use延迟对性能非常重要**
  - **Load指令占所有指令的30%以上**
  - **Load指令与使用其结果的指令需要隔几拍**
- **龙芯的9级流水线以及读Cache和tag比较分开导致load-to-use偏大：4拍**
  - **使用speculative forward和load speculation优化**
- **例：memcpy的load等待**

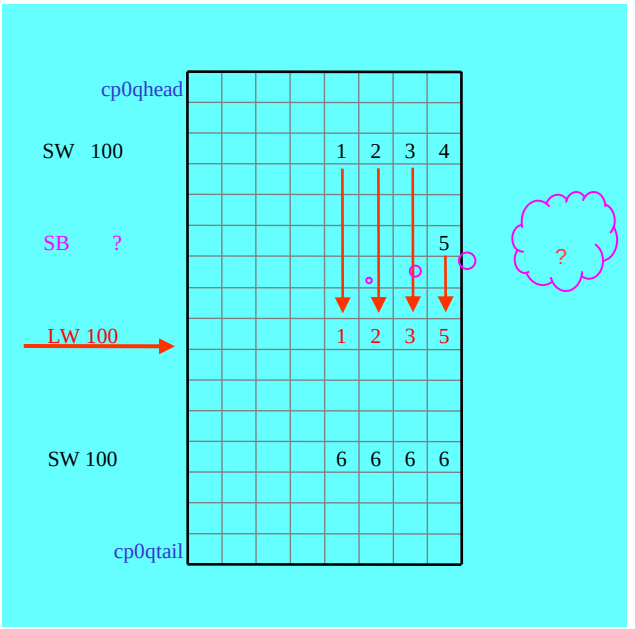
Table 5. Statistics of memory copy program segment

Program Segment	Exe. Times	Fetch/ Decoder	Register Map	Issue	Execution	Commit	Branch Miss	Cache Miss
433e58: lw $v0,0($a1)	290096	2.3	1.5	2.4	18.5	0.0	0	8609
433e5c: addiu $a2,$a2,-1	290096	2.3	1.5	2.4	1.0	17.3	0	0
433e60: addiu $a1,$a1,4	290096	2.6	1.0	2.4	1.0	16.5	0	0
433e64: sw $v0,0($v1)	290096	2.6	1.0	20.8	5.0	0.0	0	15105
433e68: bgez $a2,433e58	290096	2.6	1.0	2.9	1.0	21.9	27181	0
433e6c: addiu $v1,$v1,4	290096	2.6	1.0	3.4	1.0	21.5	0	0

降低Load-to-use延迟

- **Speculative forwarding**
  - 在读Cache阶段通知保留栈后续相关操作可以发射
  - 在tag比较阶段如果发现cache不命中，通知保留栈取消发射
- **Load speculation**
  - Cache命中的load操作必须等它前面的所有store的地址都确定后才能把值写回寄存器并传递给后面的操作（30%-40%的概率不能返回）
  - load操作Cache命中时直接返回，并在发现访存相关时取消该load及其后面的操作（<<1%的概率需要取消）





Load Speculation引起的性能提高

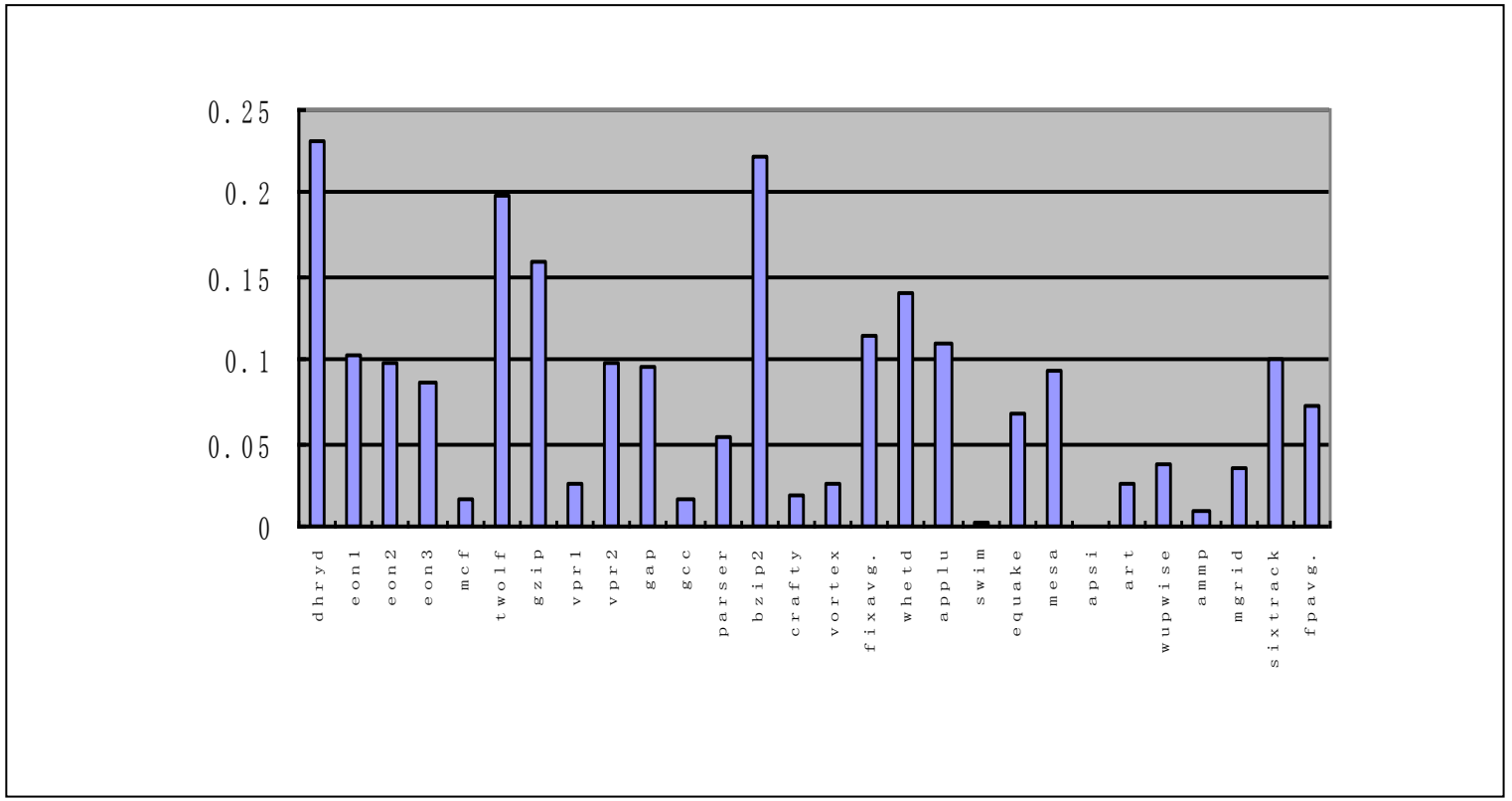


Figure 8. Performance improvement of Load Speculation Optimization

故事9: 降低Store失效开销

- **龙芯2号实现写回（Write-Back）式Cache**
  - **访问失效时，从下一级存储取回失效块再进行读写**
  - **研究表明，对于写失效，多数情况下取回的Cache块会被完全覆盖**
  - **实现Store-Fill-Buffer优化，直接在失效队列进行拼接**

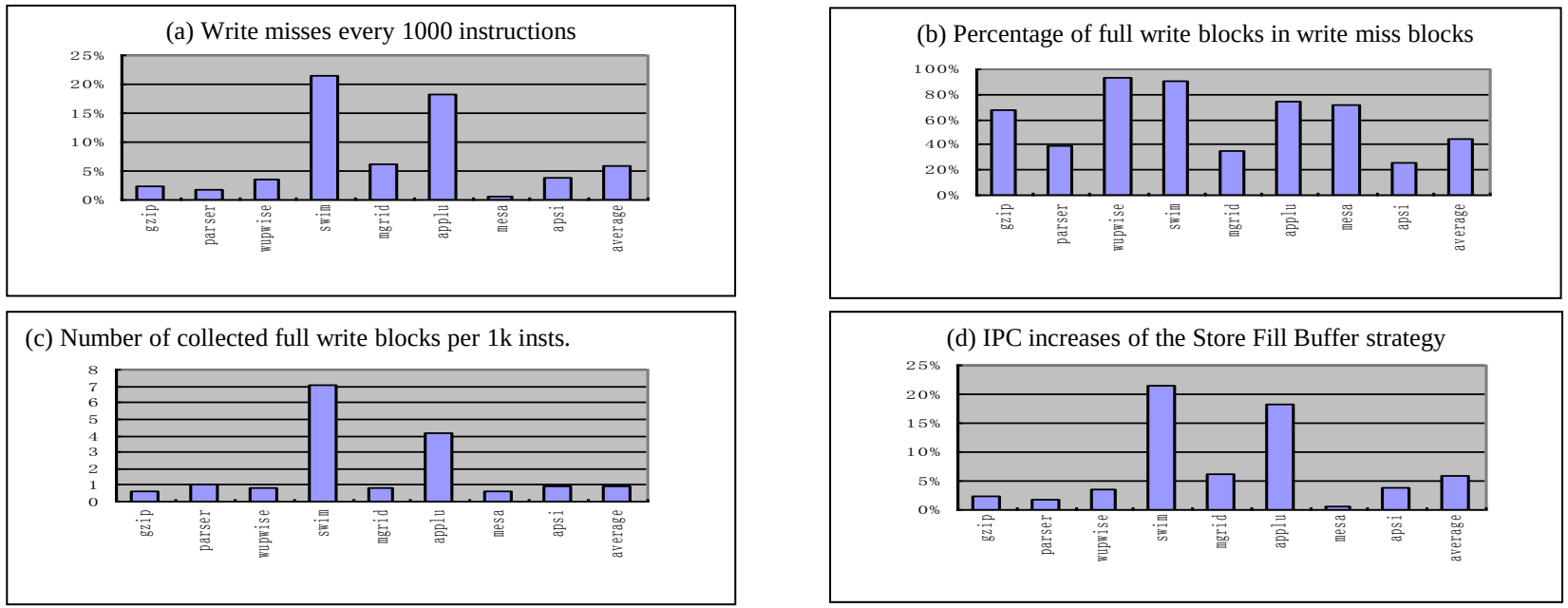


Figure 9. Store Fill Buffer Optimization

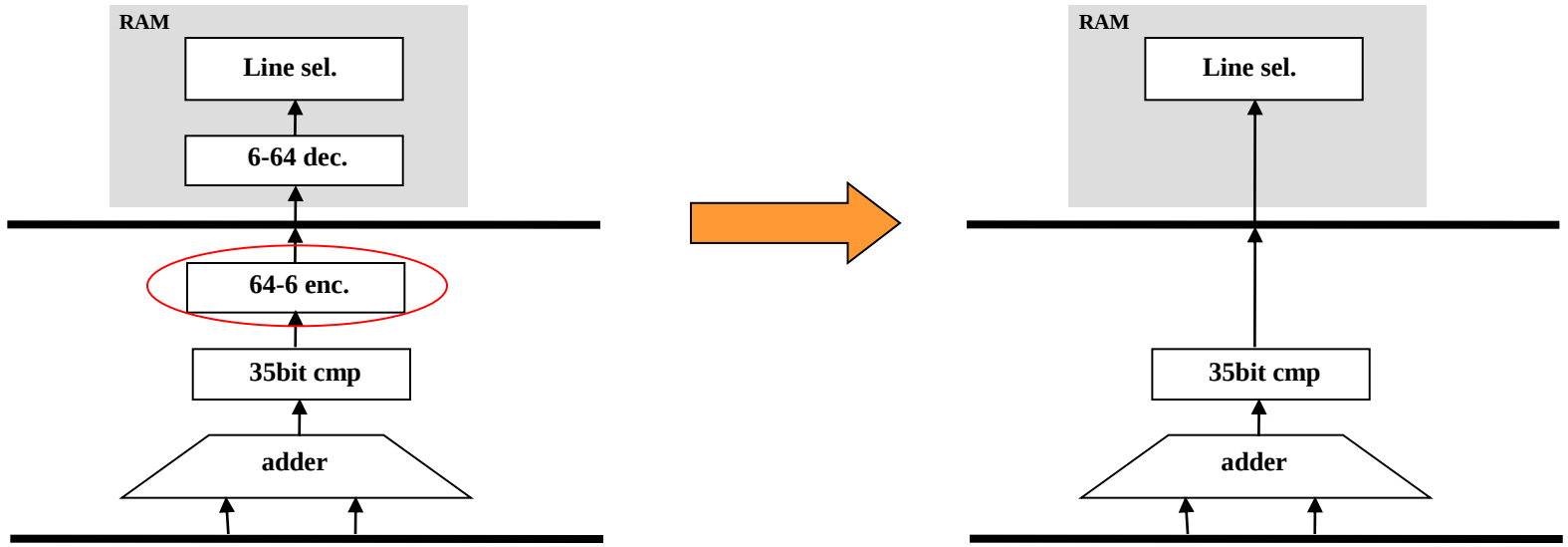
面向工艺的结构设计

皮体系结构（Pico-Architecture）设计

- 影响体系结构设计的因素
  - 工艺和应用
  - 体系结构设计要做到“上知天文（应用、OS、编译）、下知地理（电路、工艺）”
  - 工艺对处理器系统结构的影响大于应用的影响
- 纳米级工艺的体系结构设计新特点
  - 存储墙=>功耗、连线延迟
  - 纳米级的体系结构设计要充分考虑版图规划和连线延迟
  - 从**micro-architecture**到**pico-architecture**
  - **Pico-architecture**：面向物理实现的结构和组织
- 结构设计和物理设计的紧密结合和融会贯通是龙芯的重要优势和经验
  - 通过若干结构、逻辑优化，关键路径逻辑单元从龙芯**2B**的**34**级降低到龙芯**2C**的**21**级
  - 使用相同工艺，龙芯**2B**主频**250MHz**，龙芯**2C**主频**450MHz**

故事10-1：降低地址运算和TLB访问延迟

- 地址运算关键路径物理优化
  - 48位加法器、35位比较、64-6编码
  - 64-6的编码和下一级访问RAM的6-64译码“抵消”
  - 需要定制新的RAM模块



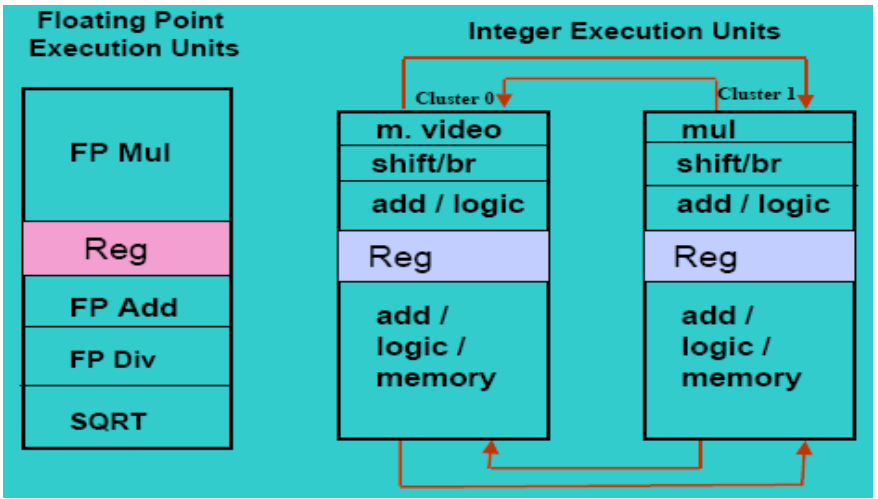
(a) With encoding and decoding

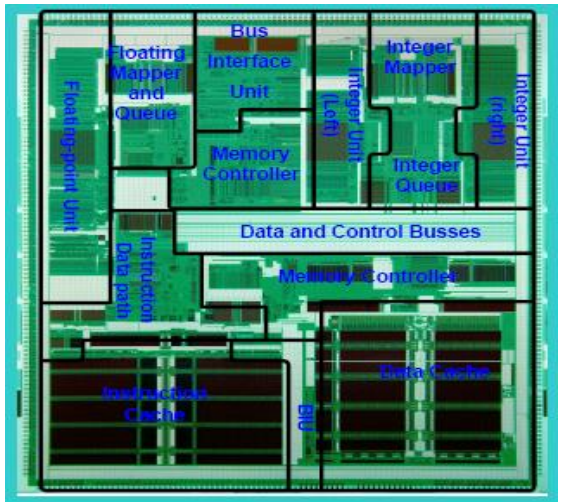
(b) Removing encoding and decoding

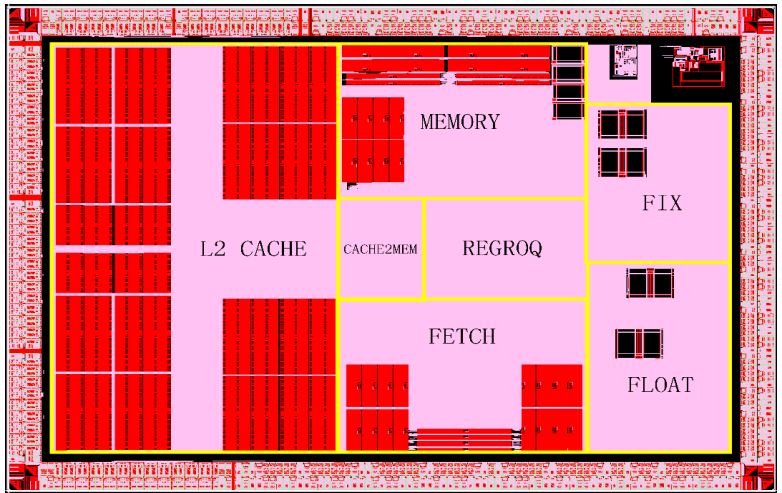
Figure 11. TLB CAM optimization

故事10-2：寄存器堆延迟优化

- 根据物理位置进行运算结果的Forward
  - 定点forward给定点，浮点forward给浮点
  - 访存forward给定点
- 按照就近的原则使用两个4w4r的寄存器堆实现4w8r寄存器堆
  - 定点：ALU1（1w3r）、ALU2（1w2r）、访存（2w3r）
  - 浮点：FALU1（1w3r）、FALU2（1w3r）、访存（2w2r）







体会

- 不要被模拟器欺骗，更不要自己骗自己
- 软硬件的融合是提高性能的关键
- 永远喂不饱的CPU
  - 带宽是冯诺依曼结构永恒的主题
- 平衡的设计至关重要
- 优化最频繁的事件，好钢用在刀刃上
  - 投资存储总有回报
- 皮体系结构设计的重要性，结构跟物理的结合

体　会

- “发现结构瓶颈、改进设计、在期待中流片、收获性能提高、发现新瓶颈”的螺旋上升的改进过程是一个充满痛苦并激动人心的过程，设计人员的经验随着设计的提高而提高
  - 结构的魅力在于统筹兼顾，质量的魅力在于持续改进
  - 好的结构设计师是用失误喂出来的
- 持续改进、“work-on-silicon”的设计态度以及对软硬件及工艺的融会贯通是一个优秀设计师的三个最重要品质

纸上得来终觉浅

绝知此事要躬行