

小容量高性能 SRAM 的设计与实现

秦海阳 李 勇 李振涛 张秋萍

(国防科学技术大学计算机学院 长沙 410073)
(563133342@qq.com)

Design and Implement of Small Capacity SRAM with High-Performance

Qin Haiyang, Li Yong, Li Zhentao, and Zhang Qiuping

(College of Computer, National University of Defense Technology, Changsha 410073)

Abstract The first level cache of microprocessor needs small capacity SRAM with high speed. Based on storage array formed by 8 tube SRAM cell, the paper builds the peripheral circuit and designs a capacity of 32×32 SRAM completely. Full-custom design of memory is compared with the memory created by Memory Compiler and the memory created by semi-custom design with the code of RTL. We find out that full-custom design of memory's performance is better than the other two on the speed, power consumption and area.

Key words 8 tube SRAM cell; circuit design; layout design; SRAM performance analysis; SRAM performance comparison

摘 要 微处理器的第一级高速缓存需要速度快的小容量 SRAM 存储器,以 8 管 SRAM 单元组成存储阵列,然后构建外围电路,设计一个容量为 32×32 的 SRAM 存储器.将定制设计的存储器与 Memory Compiler 生成的存储器和 RTL 级代码进行半定制设计的存储器进行对比.通过对比发现,定制设计存储器的性能比其他二者在速度、功耗和面积上都要好.

关键词 8 管 SRAM 单元;SRAM 电路设计;SRAM 版图设计;SRAM 性能分析;SRAM 性能比较

中图法分类号 TP302

在 DSP 芯片中,高速缓存体系结构采用两级实时高速缓存,分别是第 1 级高速缓存(L1)和第 2 级高速缓存(L2)^[1].L1 用于缓存较低字节的数据和程序指令,消除程序和数据总线对存储器资源的冲突.由于受管芯面积不能太大的限制,第 1 级存储器的容量不能太大,并要求具有很快的速度.基于 L1 在缓存系统的重要性,设计系列小容量高性能 SRAM 存储器用于满足现实需要.

设计系列不同容量的小容量 SRAM 存储器,需要掌握小容量 SRAM 存储器的通用设计方法.定制设计小容量 SRAM 存储器已经开始在 XXXX-DSP 系列芯片中应用,并且能够满足 1 GHz 的频率要求.本文以存储容量 32×32 的存储器设计为例,拓

展引申一系列小容量存储器的设计方法.

1 电路设计

全定制 SRAM 32×32 的存储器由 8 管 SRAM 单元组成存储阵列,然后在存储阵列的基础上构建外围电路而成.SRAM 设计是在满足稳定性基础上,再寻求降低时序、功耗和面积的优化设计^[2].

1.1 电路整体结构

SRAM 存储器的电路结构由门控时钟、地址锁存、读写译码、存储阵列和 IO 电路 4 部分组成,它的结构框图如图 1 所示:

收稿日期:2014-09-23

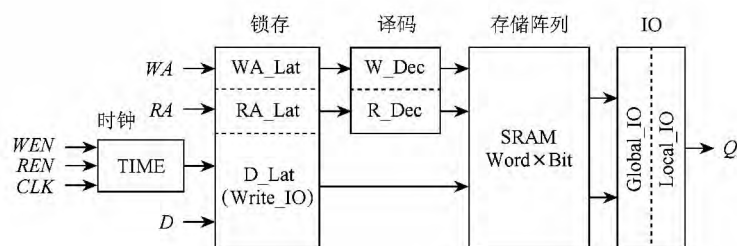
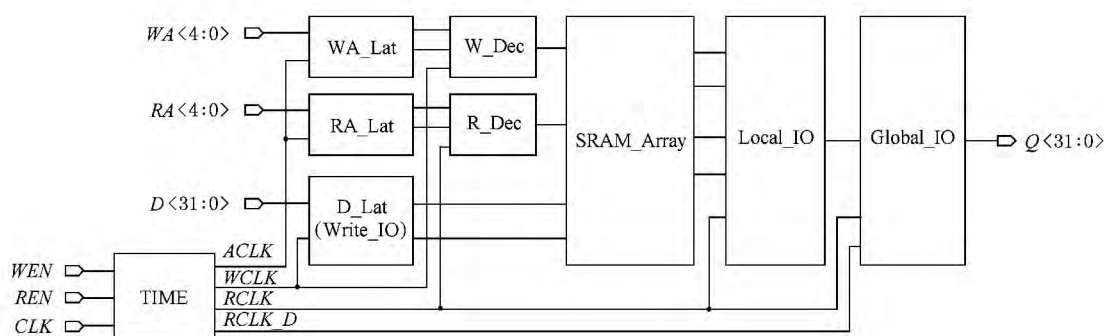


图 1 SRAM 存储器结构框图

从图 1 看出,电路的关键路径依次经由时钟模块、锁存模块、译码模块、存储阵列和 IO 模块。锁存模块对读写地址和数据进行锁存;IO 模块通过预充和下拉的方式输出读出位线值,然后对读出值进行触发锁存;读写地址分别进行译码,用以控制存储阵

列的存储单元进行定向读写操作;写入数据经由触发锁存之后,直接进入存储阵列以待存储。

按照 SRAM 存储器结构框图,设置存储容量为 32×32 的 SRAM 存储器,SRAM 32×32 存储器的电路结构如图 2 所示:

图 2 SRAM 32×32 电路结构图

从图 2 看出,TIME 为时钟模块;WA_Lat 为写地址锁存(即写 IO),RA_Lat 为读地址锁存,D_Lat 为数据锁存(即写 IO);W_Dec 为写译码,R_Dec 为读译码;SRAM_Array 为 SRAM 单元存储阵列;Local_IO 为局部 IO,Global_IO 为全局 IO。

SRAM 32×32 存储器的整体功能实现,是将各个模块的具体设计实现而成。通过电路分体结构设计,实现存储器电路的整体设计。

1.2 电路分体结构

电路分体结构由时钟模块、锁存模块、译码模块、存储阵列和 IO 模块 5 个部分组成。下面分别就这 5 个模块的设计与实现进行电路设计。

1.2.1 时钟模块

在时钟模块,通过时钟信号 CLK、写使能 WEN 和读使能 REN,产生另一个时钟信号 ACLK 和 2 个门控时钟信号 WCLK、RCLK 以及一个脉冲时钟信号 RCLK_D。时钟模块的电路设计如图 3 所示。

从图 3 看出,存在 2 个关键逻辑电路,一个是门控时钟产生电路,另一个是窄脉冲信号产生电路。读和写的门控时钟波形原理是一样的,以写门控时钟为例,WEN、CLK、WCLK 三者的波形关系如图 4 所示。

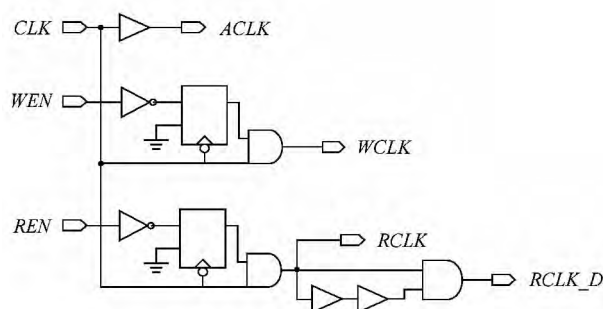


图 3 时钟模块电路

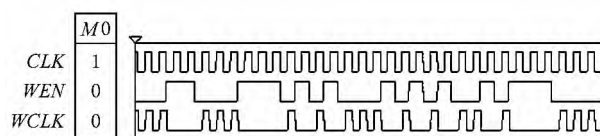


图 4 写门控时钟波形

从图 4 看出,WEN 为低电平有效,当 WEN=0,CLK=WCLK;当 WEN=1,WCLK=0。即 WEN 有效时,WCLK 进行按时钟频率进行翻转;当 WEN 无效时,WCLK 保持低电平不进行翻转。使用门控时钟可以降低时钟的不必要翻转,减少功耗损失。

窄脉冲电路由两级 buffer 和一个 and 门构成, 窄脉冲的波形产生原理如图 5 所示:

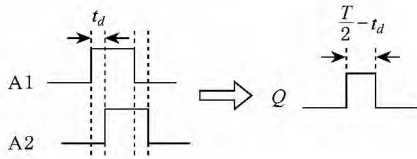


图 5 窄脉冲波形产生

从图 5 看出, $RCLK$ 先到达 $A1$ 端口, 后到达 $A2$ 端口, 两者产生延时差, 经过相与之后产生窄脉冲波形 $RCLK_D$, $RCLK_D$ 的脉宽比 $RCLK$ 短 t_d , 为 $\frac{T}{2} - t_d$.

1.2.2 锁存模块

锁存模块分为 2 部分, 分别是地址锁存和数据锁存. 数据锁存即为写 IO 电路, 由于写 IO 的功能和在路径中所处的位置, 故而将写 IO 放在这里而显得层次结构更加分明. 读写地址锁存电路设计如图 6 所示:

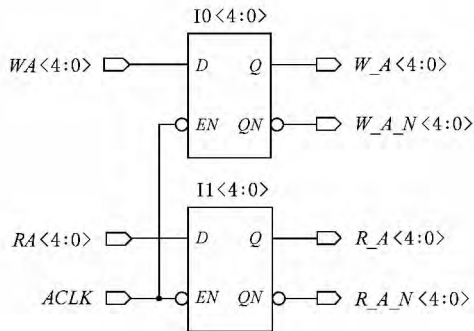


图 6 地址锁存电路

从图 6 看出, 锁存器为负锁存器, 当 $ACLK$ 为低电平, 锁存器为透明模式, 读写地址从 D 传到 Q 输出, 以及取反后传到 QN 输出; 当 $ACLK$ 为高电平, 锁存器为维持模式, 保持 Q 和 QN 处的值. 数据锁存电路如图 7 所示:

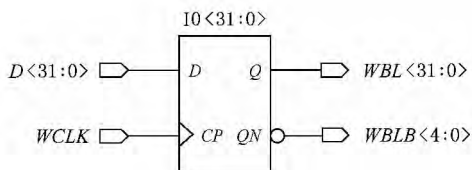


图 7 数据锁存电路

锁存单元为边缘触发器, 当 $WCLK$ 的上升沿到来之时, 触发器瞬间导通, 数据从 D 端口到达 Q 端口; 当 $WCLK$ 为高电平或者低电平, 触发器处于

维持状态, Q 和 QN 的值 WBL 和 $WBLB$ 的值保持不变, 保持的时间为一个时钟周期, 相当于写数据被寄存了一个时钟周期.

1.2.3 译码模块

在小容量存储器电路, 一般采用两级静态译码策略, 即预译码-最终译码的结构. 由于译码地址位宽是 5, 将预译码分成 3-8 译码和 2-4 译码 2 部分. 最终译码是在门控时钟的控制下对 2 部分预译码的结果再次进行译码. 译码电路结构如图 8 所示:

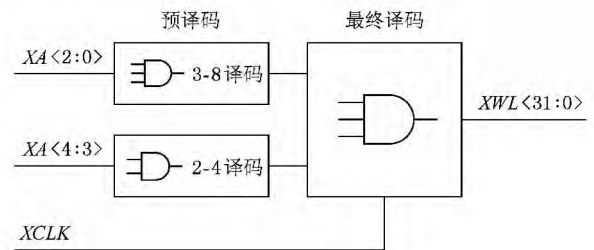


图 8 二级译码电路

如图 8 所示, 在预译码的 3-8 译码中, 有 8 个三输入与门; 在预译码的 2-4 译码中, 有 4 个二输入与门; 在最终译码有 32 个三输入与门.

1.2.4 存储模块

存储模块是存储阵列电路, 用于对数据进行缓存, 支持同时进行读写操作, 但不允许对同一行存储单元同时进行读写操作, 存储阵列电路如图 9 所示:

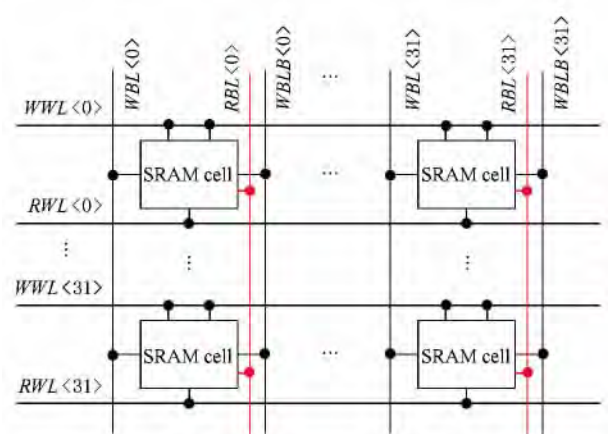


图 9 存储阵列电路

从图 9 看出, 有 32 根读字线 (RWL) 和写字线 (WWL), 32 根互补的写位线 WBL 和 $WBLB$, 以及 32 根读出位线 (RBL). 通过写字线控制数据从写位线写入存储单元, 通过读字线控制存储数值从存储单元读取到读位线. 存储单元采用 8 管 SRAM 单元, 它的电路结构如图 10 所示:

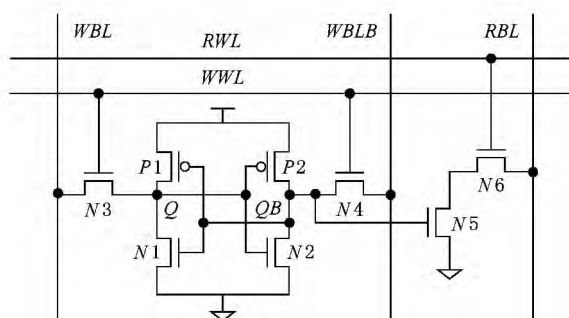


图 10 8T SRAM 单元电路

从图 10 可以看出, WWL 控制 $N3$ 和 $N4$ 管的开启, 当写字线为高时, WBL 和 $WBLB$ 的值分别写入到 Q 和 QB 。读出则是通过预充下拉的方式实现, 在进行读出操作时 RWL 为高电平, $N6$ 管导通, 同时 RBL 在读 IO 电路已被预充到 1。在进行读操作时, 当 $Q=1$, 则 $QB=0$ 使得 $N5$ 关闭, RBL 保持预充值, 即 $RBL=1$, 从而 $Q=1$ 从存储单元读出到

RBL ; 当 $Q=0$, 则 $QB=1$ 使得 $N5$ 开启, 因 $N6$ 管也开启, 可将 $RBL=1$ 下拉到 0, 从而实现 $Q=0$ 值从存储单元读出到 RBL 。

8 管 SRAM 单元与传统 6 管 SRAM 单元相比较, 在稳定性、抗噪声、漏电流和速度等方面都具有优势, 影响存储器外围电路设计, 决定存储器的整体性能^[3]。

1.2.5 读出 IO 电路

IO 电路的写入 IO 已经放在锁存模块讲解, 此处的 IO 电路主要介绍读出 IO 的设计, 读出 IO 又分为局部 IO 和全局 IO。

1) 局部 IO 电路

局部 IO 电路将有效读出位线通过预充下拉的方式进行选择输出。读 0 输出和读 1 输出分别走 2 条不同路径, 局部 IO 电路结构和传输路径波形图, 如图 11 所示:

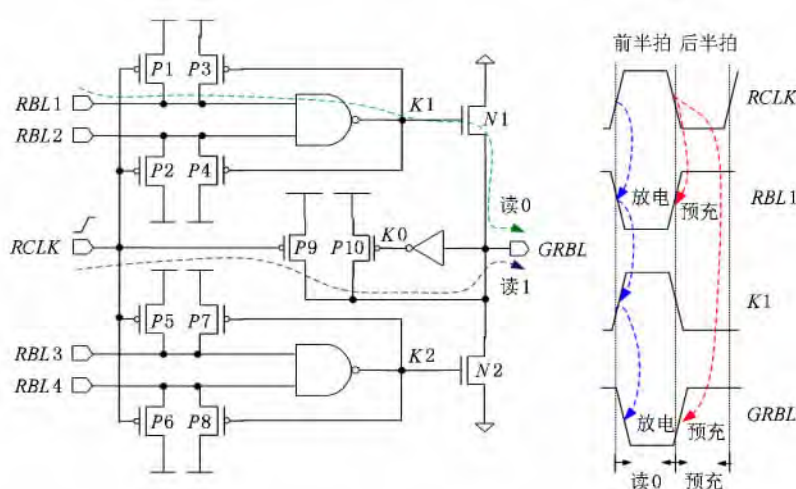


图 11 局部 IO 预充下拉电路

从图 11 可以看出, $RBL1, RBL2, RBL3$ 和 $RBL4$ 为 4 根读位线, 其中的 1 根为有效读位线。 $RCLK$ 为读门控时钟, 在前半拍将有效 RBL 值传到 $GRBL$, 在后半拍对 $RBL1, RBL2, RBL3, RBL4$ 和 $GRBL$ 进行预充。

在上一时钟周期后半拍已将 $RBL1, RBL2, RBL3$ 和 $RBL4$ 预充到 1, 在当前时钟周期的前半拍, 若有效读出值为 1, 则 4 根读出位线仍都为 1, 使得 $K1$ 和 $K2$ 的值为 0, $N1$ 和 $N2$ 管都关闭, $GRBL$ 则为预充值 1, 从而使得 $RBL=1$ 的值传到 $GRBL$ 。读 1 的路径如图 11 下方虚线箭头所示。

若有效读出值为 0, 并且该值处于 $RBL1$ 。在 $RCLK$ 的前半拍, 即 $RCLK$ 为高电平, $P1, P2, P5$ 和 $P6$ 都关闭, 存储单元 RWL 有效, 将 $RBL1$ 放电

至低电平。在 $RBL1=0$ 时, 因 $RBL2=1, RBL3=1, RBL4=1$, 所以 $K1=1, K2=0$, 此时 $N1$ 管导通, 将 $GRBL$ 下拉至低电平, 从而使得 $RBL1=0$ 的值传到 $GRBL$ 。读 0 的路径如图 11 上方虚线箭头所示。

局部 IO 电路中存在 2 处静态泄露器补偿电荷泄露电路, 如图 12 所示:

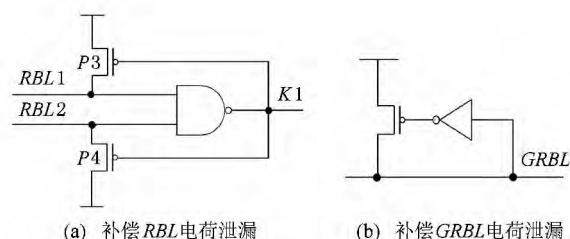


图 12 静态泄露器补偿电荷泄露电路

在读值期间,下拉路径关断时,输出节点处于高阻状态,引起漏流现象.漏电问题通过降低求值期间输出节点的输出阻抗解决.如图12(a)所示,增加2个泄漏晶体管 $P3$ 和 $P4$,图12(b)用了1个泄漏晶体管.为减少静态功耗,使泄漏器的电阻较高,尽量减少器件尺寸.图12(a)使用与非门,图12(b)使用反相器,分别驱动各自的泄漏器件,以抵抗漏电和电荷的重新分布^[4].

2) 全局 IO 电路

全局 IO 电路将读出值用窄脉冲进行触发锁存,使半个周期的读出值被触发锁存后,变成1个时钟周期的读出值进行输出.全局 IO 的电路结构如图13所示:

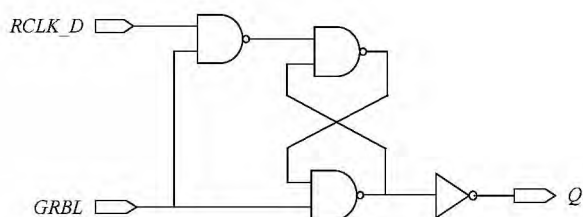


图13 SR触发锁存电路

图13为基于NAND门的SR触发器,两个与非门构成的双稳元件. $RCLK_D$ 是窄脉冲信号, $GRBL$ 为最终读出位线值.

当 $RCLK_D=1$,SR触发器处于导通状态;当 $RCLK_D=0$,SR触发器处于维持状态.经过 $RCLK_D$ 信号对触发器进行导通和保持,使半个时钟周期的 $GRBL$ 信号在 Q 端被拉长成一个时钟周期输出.SR触发器进行锁存的波形原理如图14所示:

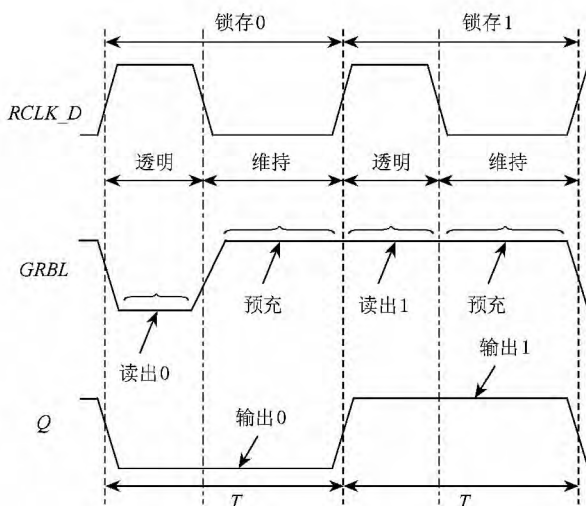


图14 SR触发器进行锁存的波形原理图

从图14看出,锁存0和锁存1的操作是通过在 $RCLK_D$ 的上升沿开始采样,在高电平阶段让 $GRBL$

信号通过,在低电平阶段对通过的信号进行维持,从而输出 Q 的脉宽被拉长到1个时钟周期.

若 $RCLK_D$ 信号比 $GRBL$ 信号来得早,在连续两拍进行锁存0操作时,便会在两个输出 Q 值的连接位置产生毛刺,毛刺产生和消除的波形原理如图15所示:

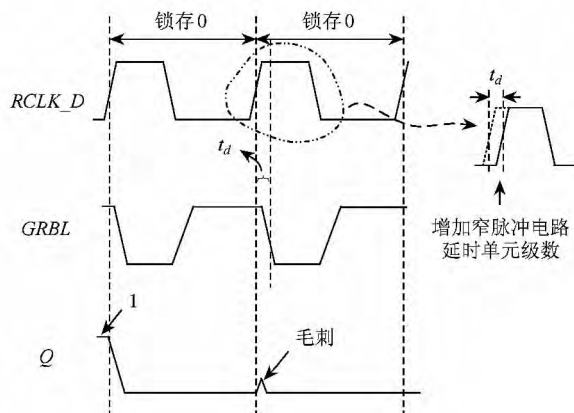


图15 毛刺产生和消除的波形原理图

毛刺的产生不仅增加了漏电流功耗,而且在不同工艺环境毛刺有被进一步放大的可能,如果毛刺足够大,还可能造成输出结果在其他地方触发锁存时有被错误采样的危险.

$RCLK_D$ 的上升沿到达时间,可以通过调整时钟模块窄脉冲产生电路的延时单元进行控制,如图15所示, $RCLK_D$ 的上升沿与 $GRBL$ 的下降沿的延时差为 t_d ,可以通过增加窄脉冲电路延时单元的逻辑级数,使 $RCLK_D$ 的上升沿迟来 t_d 个时间差,从而消除毛刺产生.

$RCLK_D$ 的脉冲宽度越小,正确锁存 $GRBL$ 值的可能性越大,在不同工艺环境锁存错误的可能性越小.然而 $RCLK_D$ 脉宽越小,窄脉冲电路延时单元越多,增加电路面积和功耗,同时还增加 Q 的输出延时.因此, $RCLK_D$ 的脉冲宽度在满足不同工艺拐角不出现毛刺的基础上,尽量加大 $RCLK_D$ 的脉冲宽度.

2 版图设计

存储器版图的整体布局布线,采用左右对称的布局规划和布线设计,布局更加合理紧密,布线有效长度更短,有助于提高版图物理实现时的速度和稳定性,有利于降低版图的面积^[5].

2.1 版图布局

将存储阵列、读出 IO 和数据锁存(写入 IO)模

块一分为二放在版图两侧,将译码、时钟、读写地址锁存模块放在版图中间,形成一种左右对称的镜像式对称。

版图布局如 16 所示,图中小黑点处于方格中的位置表示同一类方格的位置朝向对应关系;方格中的数字表示与具体某一根位线的对应关系。这种版图布局方法,可为其他系列不同容量的小容量存储器布局提供参考。

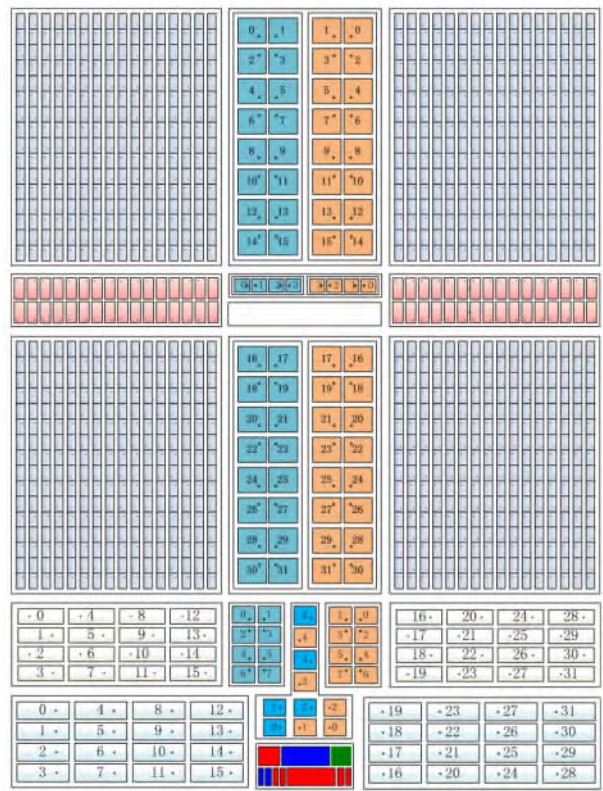


图 16 存储器版图布局

从图 16 可以看出,除了地址锁存模块是表面对称、时钟模块不对称之外,其他模块的版图都是绝对的左右对称。将时钟、译码放在中间位置,可以让产生的时钟信号、译码信号更近距离、更有效地控制两边的单元。

2.2 版图布线

版图合理的布局是为版图的布线服务,而布线相对于布局则更加复杂。版图以整体对称的方式布局,依附在其上的布线也是以对称的方式布线。

从图 17 可以看出,整个版图以对称方式布线,存储阵列布线最为密集,最为整齐;译码部分布线最为复杂,也很密集。从译码置于版图中间位置,译码出来的读写字线可以短距离快速控制存储单元的读和写。

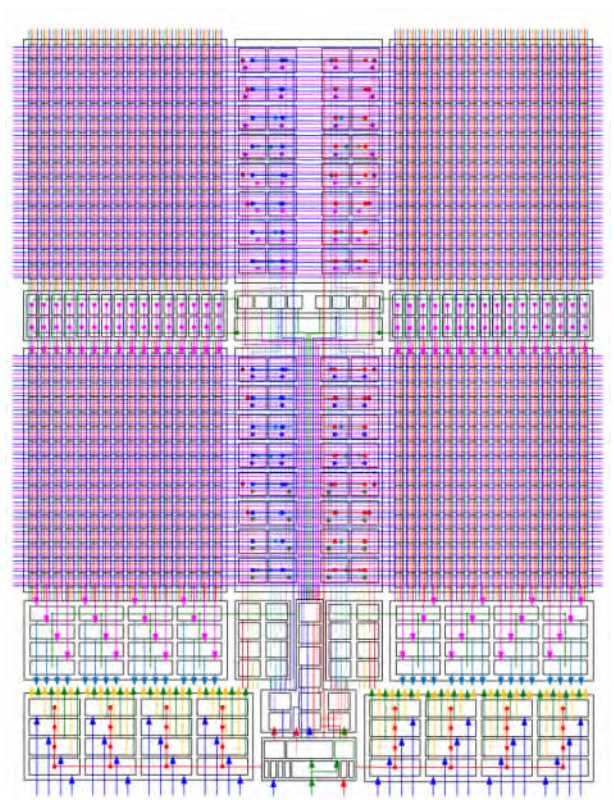


图 17 存储器版图布线

3 性能比较

使用 XA 工具进行功耗和时序测量,测得 CLK-Q 的延时和平均功耗。在 TT 工艺拐角,将定制设计的存储器与半定制设计和编译器设计的存储器,在面积、时序和功耗上分别进行数据对比,如表 1 所示:

表 1 不同版本设计比较			
参数	全定制设计	半定制设计	编译器设计
面积/ μm^2	2 330	8 951	3 856
时序/ns	0.27	0.33	0.42
功耗/mW	1.06	1.64	2.34

从表 1 可以看出,全定制存储器设计在面积、时序和功耗等方面都比其他 2 个设计性能优越。为了更加具体详细地对比性能差距,将三者制成柱状图进行对比,如图 18 所示。

从图 18 可以看出,全定制设计比半定制设计的面积小 74%,时序小 18%,功耗小 35%;全定制设计比编译器设计的面积小 40%,时序小 35%,功耗小 55%。

全定制设计比编译器设计存储器面积小 40%,

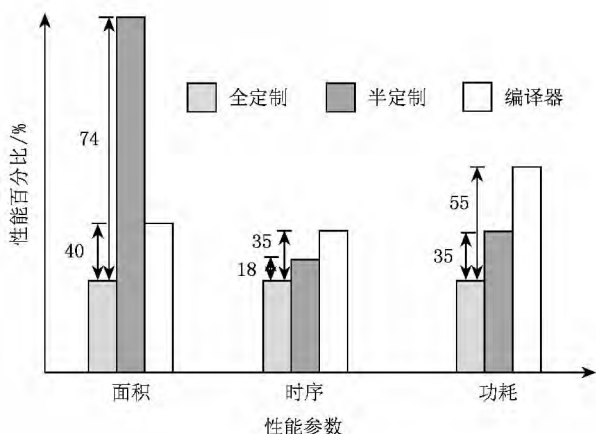


图 18 3 种存储器设计的柱状比例图

这是因为虽然全定制设计的存储阵列面积比编译器设计的要略大,但是全定制设计的 IO 面积却远小于编译器设计的 IO 面积,根本原因在于编译器采用 6 管 SRAM 单元所致. 不过编译器设计开销最小,可以快速获得性能较好的存储器设计;定制设计虽开销最大,但却可以通过获得高性能而扩大销量来分摊成本.

4 结束语

本实验设计一个存储容量为 32×32 的小容量 SRAM 存储器,通过电路设计和版图设计完成存储器的物理实现. 电路设计是以高可靠、高速度、高密度和低功耗为设计目标,主要创新点体现在门控时钟、静态译码、动态读出和 SR 触发锁存 4 个方面;版图设计以有效走线最短为设计目标,采用左右对称的布局和布线方式.

通过性能比较发现,全定制设计的存储器比半定制设计的面积小 74%,时序小 18%,功耗小 35%;

比编译器设计的面积小 40%,时序小 35%,功耗小 55%. 由此看来,全定制设计存储器在时序、功耗和面积上都比半定制设计和编译器设计的存储器要好,说明全定制存储器设计非常成功.

SRAM 32×32 存储器定制设计的成功,有效满足第一级高速缓存对存储容量为 32×32 的存储器的现实需要,也为一系列其他小容量存储器的定制设计提供设计方法,具有很重要的参考价值和借鉴意义.

参 考 文 献

- [1] 曾莹, 伍冬, 孙磊, 等. 先进半导体存储器-结构、设计与应用. 北京: 电子工业出版社, 2005
- [2] Rajesh Manapat, Manoj Roge. 移动应用中高密度低功率 1TPSRAM 的新兴架构. 中国集成电路, 2002(9): 33-35
- [3] 温亮, 李勇, 李振涛. 新型 SRAM 单元的分析与比较. 第十四届计算机工程与工艺年会. 长沙: 国防科学技术大学出版社, 2010
- [4] Jan M R, Anantha C. 数字集成电路—电路、系统与设计. 第二版. 北京: 电子工业出版社, 2004, 10
- [5] 姜岩峰. 现代集成电路版图设计. 北京: 化学工业出版社, 2009

秦海阳 男, 1986 年生, 硕士, 主要研究方向为高性能集成电路设计与实现.

李 勇 男, 1970 年生, 副研究员, 主要研究方向为微处理器和 SOC 设计.

李振涛 男, 1976 年生, 副研究员, 主要研究方向为高性能集成电路设计和 EDA 设计.

张秋萍 女, 1988 年生, 版图设计工程师, 主要研究方向存储器的电路设计和版图设计.