# 双阈值多米诺或门性能估计及其工艺浮动下的适用性分析

汪金辉[1],宫 娜[2],左 磊[1],彭晓宏[1],吴武臣[1]
(1.北京工业大学集成电路与系统研究室,北京,100124;2.纽约州立大学布法罗校区计算机学院,纽约布法罗,14260 美国)

**摘 要:** 提出了一种基于小波神经网络,估计双阈值多米诺或门的漏功耗和速度,随着扇入的增加,非线性变化的系统方法.分析表明,此方法估计误差均小于 5%,具有很高的准确性和稳定性,产生估计误差的原因为功耗比例变化和电容匹配的影响.最后,利用蒙特卡罗分析验证了此方法在工艺参数浮动下的适用性,并得出结论:在工艺参数的影响下,适用于较小扇入的漏功耗估计和较大扇入的延迟估计.

**关键词:** 小波神经网络; 双阈值多米诺或门; 延迟; 漏功耗; 蒙特卡罗分析
**中图分类号:** TN4 **文献标识码:** A **文章编号:** 0372-2112 (2010) 11-2611-05

# Performance Estimation for Dual Threshold Domino OR and Its Availability Analysis under Process Variation

WANG Jin-hui[1], GONG Na[2], ZUO Lei[1], PENG Xiao-hong[1], WU Wu-chen[1]
(1. *VLSI and System Lab, Beijing University of Technology, Beijing* 100124, *China*;
2. *Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY*, 14260, *USA*)

**Abstract:** An approach for estimating the non-line changing of the leakage power and the speed of the dual threshold domino OR gates with increasing fan-in based on Wavelet Neural Networks is proposed. Sine all of the average estimating errors are less than 5%, the estimating system has high precision. The reasons for the estimating errors are the changing ratio of the different power in the total power and the effect of capacitor match. At last, through Monte Carlo analysis, the availability of the estimating system under process variation is tested and concludes: With process variation, the estimating system is applicable to leakage power estimation with low fin-in and delay estimation with high fin-in.

**Key words:** wavelet neural networks; dual threshold domino OR gate; delay; leakage power; Monte Carlo analysis

## 1 引言

多扇入多米诺或门及相似的结构,广泛应用在微处理器的寄存器和存储器位线的设计中[1,2],随着集成电路工艺特征尺寸的减小,晶体管的阈值电压 $V_t$ 和栅氧化层的厚度 $t_{ox}$,随着电源电压的减小而按比例缩小[3,4].但是,漏功耗却随着阈值电压和栅氧化层的厚度的减小呈指数规律增加.2005 年版的 ITRS[5] 预测,亚 65nm 工艺下,漏功耗占电路总功耗的 50% 以上.同时,增加的漏功耗也会极大的影响多扇入多米诺或门的噪声容限等其他性能.因此,低漏功耗多米诺或门设计已成为当今微处理器和存储器设计的关键之一.

双阈值电压技术是业内普遍认可的抑制漏功耗的有效方法,被广泛应用于多米诺或门的设计中[6].但是,由于此技术采用速度较慢的高阈值晶体管,增加了电路的延迟,影响了电路的性能.因此,在对多米诺电路进行双阈值技术的优化之前,应对优化结果(漏功耗减小量和延迟增加量)进行估计,判断其是否满足功耗和速度的双重设计约束.在 VLSI 的设计流程中,这项工作至关重要,可以极大的减少迭代次数,从而节约设计者的时间.

由于漏功耗和速度随扇入变化的复杂性和非线性,对漏功耗和速度的估计一直困扰着 VLSI 的设计者.人工神经网络在数据处理中避免了数据分析和建模中的困难,具有抗干扰能力强、能自适应学习等优点[7].而小波神经网络以收敛速度快,较简单的拓扑结构实现函数逼近、模糊估计等特点,已在多种领域的估计研究中得到应用[8,9].因此,本文基于小波神经网络提出了一种估计双阈值多米诺或门漏功耗和速度的方法,并通过与 HSPICE 仿真结果的比较和蒙特卡罗分析,验证了其准确性及其在工艺参数浮动下的适用性.

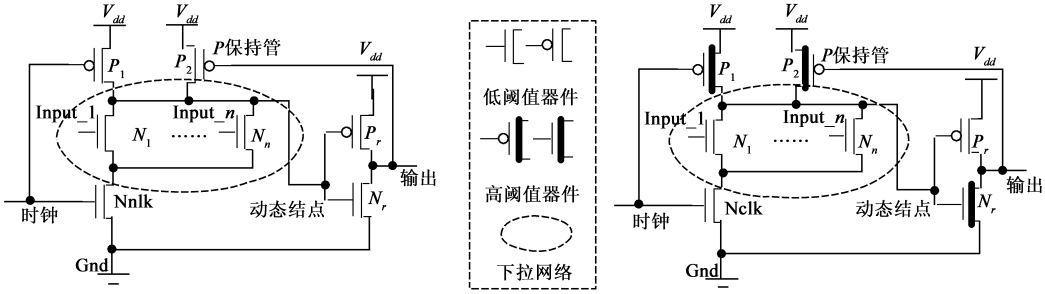
## 2 基于小波神经网络的估计系统

### 2.1 双阈值多米诺或门

由上节已知,较大的漏功耗已成为亚 65nm 多米诺或门优化的关键问题[10].为了解决这一问题,我们首先

收稿日期:2009-02-17;修回日期:2010-07-02

简要分析多米诺或门的工作原理:由图 1 所示,当时钟信号 $clock = 0$ 时,为预充阶段,$P1$ 导通,动态结点被预充到高电平 $V_{dd}$;当 $clock = 1$ 时,为求值阶段,$P1$ 管关闭,动态结点视下拉网络(PDN)有条件地放电:如果 PDN 存在从动态结点到地的直流通路,那么动态结点对地放电至低电平;否则,动态结点将借助于保持管 $P2$ 保持高电平值 $V_{dd}$,直到下一周期[11].由上述工作原理可以看出,处在关键路径的晶体管 $Nclk$, $N1\cdots\cdots Nn$, $Pr$ 决定了电路的求值速度,因此它们需采用低阈值晶体管以保证电路性能;处在非关键路径的 $P1$, $P2$, $Nr$ 对电路求值速度影响不大,可以采用高阈值晶体管以降低漏功耗.这是由于,由式(1)和式(2)可知,随着阈值电压 $V_t$ 的增加,亚阈值漏电流 $I_{sub}$ 不断减小,而速度 $v$ 明显降低[12],这即是双阈值技术的基本原理.双阈值多米诺或门电路结构如图 1($b$)所示.




($a$) 低阈值多诺米或门 ($b$) 双阈值多米诺或门

**图1 多米诺或门**

$$I_{sub} = \frac{W_{eff}}{L_{eff}} u \sqrt{\frac{q\varepsilon_{si} N_{ch}}{2\Phi_s}} V_T^2 \exp\left(\frac{V_{gs} - V_t}{nV_T}\right)\left(1 - \exp\left(-\frac{V_{ds}}{V_T}\right)\right) \tag{1}$$

$$v \propto \frac{{V_{dd}}^{0.3}\left(1 - \frac{V_t}{V_{dd}}\right)^{1.3}}{{t_{ox}}^{0.5}} \tag{2}$$

另外,下拉网络晶体管数目,即多米诺或门扇入的个数直接影响了电路的漏功耗和速度.一方面,双阈值多米诺或门的漏功耗主要由下拉网络产生,因此下拉网络中晶体管数目不同,产生的漏功耗将不同.另一方面,随着扇入的增加,地电压和动态结点之间建立了更多的电流通路,加快了求值速度,从而部分补偿了由于使用高阈值晶体管造成的速度损失.因此,在 VLSI 设计中,精确的估计双阈值多米诺或门的扇入目与漏功耗和电路求值速度之间的关系,可以帮助设计者判断其是否满足设计约束,以减小设计迭代次数和设计时间.

## 2.2 基于小波神经网络的估计系统

Zhang 于 1992 年将小波分析理论运用到神经网络中[13],通过利用非线性的小波基函数代替传统的 Sigmoid 函数,提高了神经网络的精度.

小波变换理论的实质就是将信号表示成基函数的线性组合.它通过对基函数的母函数进行伸缩和平移,得到一个小波序列:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{1-b}{a}\right)a, b \in R, a \neq 0 \tag{3}$$

其中,$\psi(t)$为母函数,$a$ 是时间轴伸缩因子,$b$ 是时间平移因子.则小波神经网络的输出为:

$$f(x) = \sum_{o=1}^{h}\sum_{j=1}^{m} \omega_o \frac{1}{\sqrt{|a_o|}}\psi_{jo}\left(\frac{\sum_{i=1}^{n} x_{ij} - b_o}{a_o}\right) \tag{4}$$

其中,$\omega_0(0 = 1,2,\cdots,h)$表示隐层第 0 个单元的输出权值;$\psi_{i0}$表示第 $j$ 个输入样本在第 0 个隐层单元基函数作用下的输出值.式(2)表明,函数 $f(x)$可以用小波神经网络来逼近.显然,网络中待训练的参数有三个:输出权值 $\omega$,伸缩因子 $a$ 和平移因子 $b$.

定义 $D = \{(y_i, x_i) \mid i = 1,2\cdots, N\}$为训练样本集,$N$ 是训练集的样本数,对于漏功耗和速度估计系统,输入为从 1 到 150 之间的扇入个数,初始输入为 1,5 至 150 之间的输入按步长 5,依次递增.小波基函数采用一维小波函数:

$$\psi(x) = \cos(1.75x)\mathrm{e}^{-x^2/2} \tag{5}$$

取下列指标作为适应度函数:

$$J = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{m}(y_{j,i}^d - y_{j,i})^2 \tag{6}$$

$y_{j,i}^d$是第 $i$ 个样本的第 $j$ 个网络输出节点的理想输出值;$y_{j,i}$是第 $i$ 个样本的第 $j$ 个网络输出节点的实际输出值.

则算法的基本步骤为:

(1)初始化:确定网络需要优化的三个参数,即输出层权值 $\omega$、伸缩因子 $a$、平移因子 $b$,初始化为(0,1)之间的随机数,

(2)适应度评价:根据式(6)和训练样本集,评价每一个粒子的适应度值.

(3)粒子更新:为了最小化适应度,用以下算法对权值 $\omega$、伸缩因子 $a$、平移因子 $b$ 进行更新.

$$\omega_j(k)=\omega_j(k-1)+\eta(y^d(k)-y(k))h_j+\alpha(\omega_j(k-1)-\omega_j(k-2)) \tag{7}$$

$$\Delta a_j=a_j(y^d(k)-y(k))\omega_j h_j\frac{(\sum_{i=1}^{n}x_{ij}-b_j)^2}{a_j^3} \tag{8}$$

$$a_j(k)=a_j(k-1)+\eta\Delta a_j+\alpha(a_j(k-1)-a_j(k-2)) \tag{9}$$

$$\Delta b_j(k)=(y^d(k)-y(k))\omega_j\frac{\sum_{i=1}^{n}x_{ij}-b_j}{a_j^2} \tag{10}$$

$$b_j(k)=b_j(k-1)+\eta\Delta b_j+\alpha(b_j(k-1)-b_j(k-2)) \tag{11}$$

$$h_j=\frac{1}{\sqrt{|a_j|}}\psi_j\left(\frac{\sum_{i=1}^{n}x_{ij}-b_j}{a_j}\right)(j=1,2,\cdots,m) \tag{12}$$

这里取 $\eta=0.01$,$\alpha=0.05$,$j$ 为隐层单元数.

(4)输出判断:如果当前的迭代次数达到了预先设定的最大次数或最小目标误差,则停止迭代,输出最优解,否则转到第 2 步.

## 3　仿真结果和分析

如 2.1 节所述,基于小波神经网络的估计系统的输入为双阈值多米诺或门的扇入个数.与低阈值多米诺门相比,双阈值多米诺门漏功耗的减小量和延迟的增加量作为估计系统的输出.利用 HSPICE 工具,基于 45nm 工艺 BSIM4 模型的不同扇入多米诺门的 HSPICE 仿真结果,作为训练样本和测试样本.仿真中所有多米诺门处在漏功耗最小的 CHIH 状态[14](时钟信号为 1,所有输入信号全为 1),器件参数如表 1 所示.为了验证估计系统的有效性,测试样本选择输入分别为 2,4,8,16,32,48,64,96,128 的多米诺或门,因为这些典型的或门经常使用在实际的电路中,非常具有代表性.估计系统采用较为常用的 Morlet 小波作为网络隐含层的激活函数,网络训练的误差精度为 $\varepsilon=0.00001$,最大迭代次数选为 2500 次,此小波神经网络模型算法采用 MATLAB 编程实现,实现了 0.00001 的精度.

**表 1　器件参数**

| 特征尺寸 | 45nm | |
|---|---|---|
| | 低阈值 | 高阈值 |
| NMOS 管 | 0.22V | 0.35V |
| PMOS 管 | -0.22V | -0.35V |
| 电源电压 | 0.8V | 0.8V |

测试误差如表 2 所示,从表中可以看出,估计系统具有很高的准确性和稳定性,所有漏功耗估计误差和延迟估计误差都小于 5%,所以此估计系统可以内嵌在 EDA 软件中,作为电路设计前的预测工具.

**表 2　测试误差**

| Input | 2 | 4 | 8 | 16 | 32 | 48 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|---|---|---|
| *LR*/% | +3.80 | +1.99 | -1.26 | -2.11 | -1.77 | -2.81 | -2.96 | -1.87 | -1.81 |
| *DI*/% | +0.02 | -1.51 | +2.30 | +0.28 | -3.92 | +1.39 | +2.96 | +2.46 | -3.77 |

*LR*:漏功耗减小的百分比. *DI*:延迟增加的百分比

图 2 示出了基于小波神经网络的系统估计曲线.从图中可以看出,当扇入数量小于 8 时,功耗减小曲线位于估计曲线之上,而当扇入大于或等于 8 时,功耗减小曲线位于估计曲线之下.这是因为在纳米尺寸下,多米诺或门的漏功耗由栅极漏功耗(栅极漏功耗由栅极漏电流 $I_{gate}$产生)和亚阈值漏功耗两部分组成[15,16].当工艺尺寸大于 90nm 时,CMOS 工艺的栅氧化层厚度 $t_{ox}$大于 20Å,栅极漏功耗可以忽略.随着工艺技术的改进,特征尺寸每减小一次,亚阈值漏功耗将增加三到五倍,而栅极漏电流会增加一个数量级,电路中栅极漏功耗的比例逐渐增大,当工艺尺寸减小到 45nm 时,栅极漏功耗已逐渐赶上亚阈值漏功耗,而成为功耗的不可忽视的一部分.而且对于多米诺或门,由于下拉网络中的晶体管为并行的拓扑结构,没有堆栈效应的作用,栅极漏功耗的产生依次累加,所以,随着扇入的增多,栅极漏功耗在总功耗中的比例越来越大[17,18].但是,双阈值技术主要用来降低亚阈值漏功耗,对栅极漏功耗几乎没有作用.因此,当扇入个数小于 8 时,漏功耗以亚阈值漏功耗为主,双阈值技术的作用较为明显,所以测试曲线显示,漏功耗的减小百分比很大,估计曲线上的值偏小.当扇入个数大于 8 时,由于忽略了此时占总的漏功耗比例较大的栅极漏功耗,高估了双阈值技术的作用,估计曲线上的值偏大,估计曲线位于测试曲线之上.

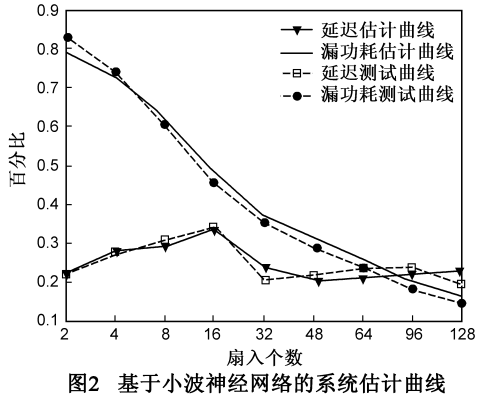



**图2　基于小波神经网络的系统估计曲线**

如 2.1 节所述,扇入个数极大影响了多米诺或门的延迟.这一延迟主要由时钟管 $P1$ 的电容,保持管 $P2$ 的电容和下拉网络的总电容 $P_{下拉}$决定.因为在电路从预充到求值的转换阶段,存在竞争电流,所以电容值 $P1+P2$ 与 $P_{下拉}$的关系决定了电路速度的快慢. $P1+P2$ 既需足够大,满足预充的要求,又需限制在一定范围内,从而限制竞争电流.当输入为 16 时,$P1+P2$ 与 $P_{下拉}$匹配最好,多米诺或门的速度最快,所以,在图 2 中可以看出,双阈值技术对 16 扇入多米诺或门的影响最大,即延迟的增加最明显.对于其他大于 16 扇入或小于 16 扇入的多米诺或门,双阈值技术对延迟的影响均明显减小.

## 4 工艺浮动下的适用性分析

在纳米级 CMOS 工艺中,制造栅长 40nm 左右的晶体管,光刻与蚀刻等环节的工艺控制是急需解决的新难题[19],这是因为工艺控制的难度带来了严重的栅长($L$)、栅氧化层厚度($t_{ox}$)和沟道掺杂浓度($N_{ch}$)等重要参数的随机浮动,这种浮动既发生在不同的晶圆之间,也发生在同一晶圆的晶体管之间.由于多米诺或门的漏功耗和延迟对这些参数具有强烈的依赖性,因此参数浮动带来了漏功耗和延迟的不均一性,这将对估计系统在工艺参数浮动下的有效性产生了一定的影响.因此,分析估计系统在工艺浮动影响下的适用性,成为又一关键问题.本文利用蒙特卡罗(Monte Carlo)仿真方法,考虑栅长($L$)、栅氧化层厚度($t_{ox}$)和沟道掺杂浓度($N_{ch}$)等重要参数的随机浮动,选取样本中最大输入 128 和最小输入 2 的多米诺或门进行了研究.仿真过程中,假设每个参数均服从高斯分布,参数 $3\sigma$ 设置为 10%,仿真次数为 1000 次[20].

图 3 示出了工艺浮动影响下,2 输入和 128 输入的多米诺或门的漏功耗和延迟分布曲线,从图中可以看出,曲线均近似为正态分布.为了进一步分析参数浮动的影响,定义了参数 $P$ = 均值/方差[18],显然,$P$ 值越小,工艺浮动影响越明显.仿真结果由表 3 示出,从表中可以看出,输入越大,漏功耗的 $P$ 值越小,漏功耗不均一性越大:输入为 2 时,漏功耗估计误差几乎不变,输入为 128 时,漏功耗估计误差为 16.99%,比原来的 1.81% 扩大了 9 倍.而输入越小,延迟的 $P$ 值越小,延迟的不均一性越大:输入为 2 时,延迟估计误差为 9.37%,是原估计误差 0.02% 的 460 倍;输入为 128 时,延迟估计误差为 5.62%,仅是原来 3.77% 的 1.5 倍.因此,可以看出,此估计系统在工艺浮动影响下,适用于较小输入的漏功耗估计和较大输入的延迟估计.

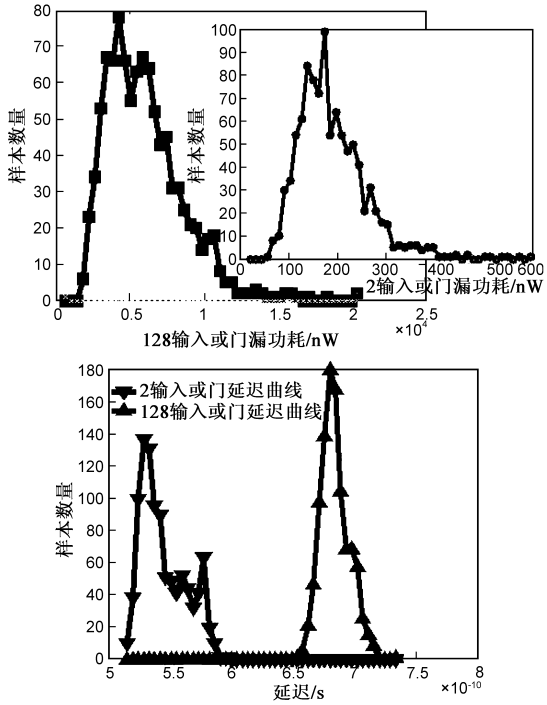



图3 2输入和128输入多米诺或门功耗、延迟分布曲线

表 3 工艺浮动影响下的估计误差

| 输入 | 2 输入或门 | | 128 输入或门 | |
|---|---|---|---|---|
| | 漏功耗/μW | 延迟/ns | 漏功耗/μW | 延迟/ns |
| 均值 | 195.1 | 54.65 | 6.185 | 68.59 |
| 方差 | 72.28 | 0.184 | 2.637 | 0.115 |
| $P$ | 2.69 | 297 | 2.35 | 596 |
| 变化量 | -79.7% | +31.3% | -0.87% | +17.3% |
| 估计误差 | 3.81% | 9.37% | 16.99% | 5.62% |

## 5 结论

基于小波神经网络,提出了一种估计 45nm 工艺下双阈值多米诺或门漏功耗和速度的具有快速收敛性和高准确性的系统方法.通过学习双阈值技术对电路漏功耗减少量和延迟增加量的影响,此系统成功估计了不同扇入多米诺或门功耗和延迟的非线性变化.通过 HSPICE 仿真验证,估计系统的准确性高达 95%.本文进一步分析了产生估计误差的原因,即系统忽略了栅极漏功耗,而随工艺尺寸的减小和扇入个数的增加,栅极漏功耗已成为漏功耗的主要组成部分.同时,本文分析了双阈值技术对电路延迟的影响,当扇入为 16 时,电路中下拉网络的电容与 P 型时钟管和 P 型保持管的电容匹配最好,双阈值技术对电路速度影响最大.最后,通过蒙特卡罗分析,本文研究了工艺参数浮动对估计系统的影响,并得出结论,系统适用于较小输入的漏功耗估计和较大输入的延迟估计.此外,通过应用、引申和扩展该估计系统的研究方法,可以建立其他估计应用系统,预测不同低功耗技术对其他逻辑门和逻辑组合门的影响,从而大大减小低功耗 VLSI 设计周期.

**参考文献:**

[1] Rusu S, Singer G. The First IA-64 Microprocessor[J]. IEEE Journal of Solid-State Circuits, 2000, 35(11): 1539 - 1544.

[2] Chatterjee B, Sachdev M, Krishnnamurthy R. Designing leakage tolerant, low power wide-OR dominos for sub-130 nm CMOS technologies [J]. Microelectronics Journal, 2005, 36(6): 801 - 809.

[3] Liu Z, Kursun V. Leakage power characteristics of dynamic circuits in nanometer CMOS technologies [J]. IEEE Transactions on Circuits and Systems II, 2006, 53(8): 692 - 696.

[4] Wang J H, et al. Low power wide dominos design in sub-65nm CMOS technologies [A]. 8th International Conference on Sol-

id-State and Integrated Circuit Technology [C]. New York: IEEE Press, 2006. 1864 – 1866.
[5] International Technology Roadmap for Semiconductors [EB/OL]. http://public.itrs.net/html, 2005.
[6] Kuroda T, et al. A 0.9 V 150 MHz 10 mW 4 mm2 2-D discrete cosine transform core processor with variable-threshold-voltage scheme [A]. Digest of Technical Papers, 43rd ISSCC [C]. New York: IEEE Press, 1996. 1770 – 1779.
[7] Luo Z, Shi Z. Wavelet neural network method for fault diagnosis of push-pull circuits [A]. 2005 International Conference on Machine Learning and Cybernetics [C]. New York: IEEE Press, 2005. 3327 – 3332.
[8] Aminian F, et al. Analog Fault Diagnosis of Actual Circuits using Neural Networks [J]. IEEE Transaction on Instrumentation and Measurement, 2002, 51(3): 544 – 550.
[9] Anant O, El-Hawary M E. Wavelet neural network based short term load forecasting of electric power system commercial load [A]. 1999 IEEE Canadian Conference on Electrical and Computer Engineering [C]. New York: IEEE Press, 1999. 1223 – 1228.
[10] Guo B Z, Gong N, Wang J H. Designing leakage-tolerant and noise-immune enhanced low power wide OR dominos in sub-70nm CMOS technologies [J]. Journal of semiconductors, 2006, 5(5): 804 – 811.
[11] Chin P, Zukowski C A, Gristede G, Kosonocky S V. Characterization of logic circuit techniques and optimization for high-leakage CMOS technologies [J]. The VLSI journal, 2005, 38(3): 491 – 504.
[12] Taur Y, Ning T H. Fundamentals of modern VLSI devices [M]. Cambridge University Press, 1998. 71 – 72.
[13] Zhang Q H, Benveniste A. Wavelet network [J]. IEEE Transaction on Neural Network. 1992, 3(6): 889 – 898.
[14] Kao J T, Chandrakasan A P. Dual-threshold voltage techniques for low power digital circuits [J]. IEEE Journal of Solid-State Circuits, 2000, 35(7): 1009 – 1018.
[15] Wang J H, et al. Low power and high performance zipper domino circuits with charge recycle path [A]. 9th International Conference on Solid-State and Integrated Circuit Technology [C]. New York: IEEE Press, 2008. 2172 – 2175.
[16] Gong N, Guo B Z, Lou J Z, Wang J H. Analysis and optimization of leakage current characteristics in sub-65nm dual Vt footed domino circuits [J]. Microelectronics Journal. 2008, 39(9): 1149 – 1155.
[17] 汪金辉,等.45nm 低功耗高性能 Zipper CMOS 多米诺全加器设计[J].电子学报.2009,37(2):266 – 271.
Wang J H, et al. Low power and high performance zipper CMOS domino full-adder design in 45nm technology [J]. Acta Electronical Sinica, 2009, 37(2): 266 – 271. (in Chinese)
[18] 王伟,等.基于测试向量中不确定位的漏电流优化技术[J].电子学报.2006,34(2):282 – 286.
Wang W, et al. Techniques of leakage current optimization based on don't care bits in test vectors [J]. Acta Electronical Sinica, 2006, 34(2): 282 – 286. (in Chinese)
[19] Meng K, Joseph R. Process variation aware cache leakage management [A]. International Conference on Low Power Electronics and Design [C]. New York: IEEE Press, 2006. 262 – 267.
[20] Liu Z, Kursun V. Leakage current starved domino logic [A]. 16th ACM Great Lakes symposium on VLSI [C]. New York: ACM Press, 2006. 428 – 433.

**作者简介:**

**汪金辉**　男,1981 年生于河北省唐山,现为北京工业大学集成电路与系统研究室博士研究生,主要研究方向:低功耗数字集成电路设计.
E-mail: wangjinhui888@yahoo.com.cn

**宫　娜**　女,1982 年生于河北景县,现为美国纽约州立大学布法罗校区计算机学院博士研究生,主要研究方向:高性能集成电路设计.
E-mail: gongna_china@yahoo.com.cn