به نام خدا

عنوان :

Speed and Power Scaling of SRAM’s

نویسنده :

امیرعلی بلورچیان

کلمات کلیدی : CMOS - SRAM – VLSI - سرعت – توان مصرفی

چکیده : در این مقاله پس از معرفی ساختار کلی SRAM ها به بررسی عوامل مهم در میزان سرعت و توان مصرفی آنها می پردازیم.

معرفی :

SRAM های پربازده و با چگالی زیاد نقش مهمی در ساختار های حافظه سیستم های محاسباتی نوین ایفا میکنند. در این مقاله قصد داریم تا تاخیر و توان SRAM ها را بر حسب تکنولوژی ساخت و اندازه مقایسه کنیم. ساخت و طراحی SRAM ها نیاز به یک حالت بالانس مابین تاخیر ، توان و سطح مقطع مورد نیاز ، دارد .در واقع برای طراحی SRAM ها می بایست داد و ستدی ما بین فاکتور های مهم گفته شده صورت داد .آنالیز مدارات گسترده SRAM با استفاده از نرم افزار های شبه سازی مانند SPICE بسیار زمان بر خواهد بود لذا از آنالیز های تقریبی و خلاصه شده استفاده میکنیم. مدل های آنالیزی و تحلیلی بسیار زیادی برای فاکتور های مورد نظر وجود دارد که می توان از آنها استفاده کرد.

ما در ابتدا ساختار کلی SRAM هارا مطالعه خواهیم کرد و سپس به تاثیر فاکتور های فوق در عملکرد آنها خواهیم پرداخت.

ساختار کلی SRAM

شکل شماره 1 ساختار تیپیکال SRAM را نشان می دهد. مسیر قابل دسترسی به SRAM ، به دو جزء قابل تقسیم می باشد : یکی دکودر می باشد که قسمتی از آدرس ورودی به word line بوده و دیگری MUX خروجی است که جزعی از سلول ها در خروجی است .

در این مقاله ما بر روی زمان خواندن اطلاعات در SRAM توجه بیشتری خواهیم داشت زیرا حساسیت بسیار بالایی دارند.

برای توانایی خواندن ورودی می بایستی کد گشایی شود تا بتواند خط word خاصی را فعال نماید. همانطور که در شکل 1 دیده میشود دیکودر معمولا خطوط word مجزایی را (تقسیم شده) به کار می برد به شکلی که قسمتی از آدرس برای فعال کردن خطوط عمودی به کار برده می شود و بقیه آدرس برای خطوط افقی به کار برده میشود. فصل مشترک این دو local word line گفته میشود. سلول هایی که به این word line متصل هستند اطلاعات خود را به bit line ها انتقال می دهند . اطلاعات از زیر بخش bit line ها از طریف MUX ستون ها به تقویت کننده ها فرستاده میشوند و از این طریق این اطلاعات تقویت شده و راهی خطوط دیتا میشوند .این سیگنال ها مجددا توسط تقویت کننده های عمومی تقویت شده و سپس از ساختار خارج میشوند .

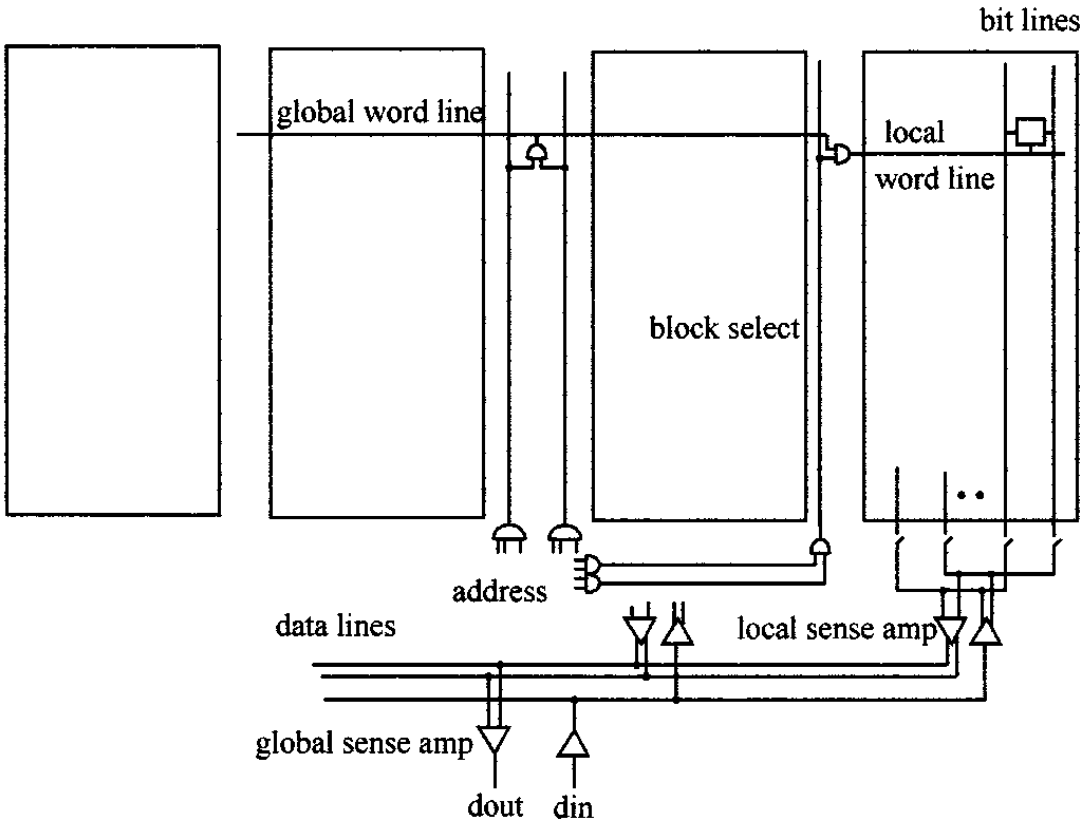
پراکندگی انرژی در SRAM ها 3 جزء کلی دارد :

1- انرژی دینامیک برای سویچ کردن ظرفیت خازنی در دکودر ها ، bit lines، خطوط دیتا و سایر سیگنال های کنترلی در ساختار.

2- انرژی آمپلیفایر های سنس شده.

3- انرژی اتلاف شده به دلیل جریان leakage.

به طور تیپیکال ساختار کلی به چند زیر ساختار با اندازه های مسای تقسیم میشود . هر کدام از این زیر ساختار ها مقداری از accessed word را در خود ذخیره میکند که به آن sub word گفته میشود .



شکل 1 : SRAM access path

این macro ها را میتوان به حالت یک RAM مجزا تصورنمود وتنها می بایست که قسمت دیکودر را به طور مشترک استفاده کنند.

مدلسازی SRAM :

برای بررسی فضای طراحی SRAM های بزرگ ما چند فرض اولیه درباره ویژگی های طراحی انجام می دهیم. ما فرض های اصلی در قسمت بعدی را مشخص می کنیم . لیستی از تمام فرض ها در پیوست موجود می باشد. سپس ما مدل های تحلیلی برای تاخیر، توان و سطح مقطع مورد نیاز اجزای SRAM را گسترش داده و با شبیه سازی مداری HSPICE مقایسه می کنیم.

این مدل ها برای بررسی کارایی رنج وسیعی از SRAM ها در اندازه های متفاوت و تکنولوژی های مختلف برای دستیابی به بهترین وضعیت از لحاظ اندازه و پیکربندی مورد استفاده قرار می گیردکه در قسمت بعدی مورد بحث قرار گرفته است.

فرضیات :

تکنولوژی اصلی که برای این تحلیل ها بکار می رود پروسه CMOS 0.25 µm می باشد که جزئیات مناسب در جدول I نشان داده شده است. پروسه مناسب که مستقل از واحد طول نامیده می شود برای تشریح پارامترهای هندسی در این مقاله بکار رفته است که برابر با نصف مینیمم اندازه در هر تکنولوژی می باشد. با فرض اینکه سرعت تمامی دستگاهها وابعاد آنها رابطه خطی با سایز و اندازه ی طرح دارند.

ابعاد منبع و ابعاد سیم بندی از مس ودر مقیاس 0.18 µm به بعد می باشد. ویژگی اصلی که برای تحلیل چهار حالت مختلف در جدول 2 نشان داده شده است.

TABLE I
FEATURES OF THE BASE 0.25-μm TECHNOLOGY

Parameter	Value	Comments
C_g	1.29 fF/µm	Unit gate capacitance of an inverter (p-width = 2 x n-width)
C_j	1.72 fF/µm	Unit parasitic junction capacitance of an inverter
R_g	4.4 KΩ-µm	Unit output Resistance of an inverter
τ_{fo4}	90pS@2.5V	Delay of an inverter driving a load four times its size.
M1 width	0.45µm	Minimum width of metal 1
M1 aspect ratio	1.5	Aspect ratio (height/width) of metal 1 cross section.
M1 cap	0.21 fF/µm	Capacitance of metal1
M1 res	0.1 Ω/µm	Resistance of metal1, (aluminum)

TABLE II
TECHNOLOGY SCALING OF SOME PARAMETERS

L (μm)	C_g (fF/μm)	τ_{fo4} (pS)	Supply (V)	M1 min. Width (μm)	M1 aspect ratio	M1 cap (fF/μm)	M1 res, type (Ω/μm)
0.25	1.29	90	2.5	0.45	1.5	0.21	0.1, al
0.18	1.29	65	1.8	0.324	1.77	0.23	0.12, cu
0.10	1.29	36	1.2	0.18	2.37	0.29	0.29, cu
0.07	1.29	25	0.9	0.126	2.83	0.33	0.49, cu

برای حذف خطا، 50 mV ضریب تطابق برایthresholds ورودی های تفاضلی در تقویت کننده حسی صرف نظر از پروسه تولید در نظر می گیریم.

تاخیر و انرژی هسته RAM را مدل کرده و از بار خارجی به علت داشتن امپدانس ثابت داخلی صرفنظر می کنیم . با فرض اینکه برای تمامی گیتهای RAM طرح مداری استاتیک برای مشخص کردن کاربردهای آن بکار می رود.تاخیر در گیت pMOS به اندازه ی تاخیر در گیت nMOS می باشد بنابراین می توان گیت را به عنوان پارامتر تک سایز مورد بررسی قرار داد. از انجا که SRAM های پر سرعت ،سایز ترانزیستور را درگیت دکودر برای بهبود مسیر بحرانی تغییری دهند لذا ما تاثیر آنرا بر روی آنالیزهای تاخیر مشخص خواهیم کرد.

تنوع زیادی در مدارهایی که برای تقویت کننده حسی بکار می رود وجود دارد. در این مقاله ما از تقویت کننده طرح latch شامل یک جفت گین cross-coupled که با کلاک فعال می شوند،استفاده می کنیم . در این ساختار تاخیر تقویت کننده متناسب با لگاریتم ولتاژ مورد نیاز می باشد.

اگر زمان بندی کلاک نمونه برداری به خوبی کنترل شود سرعت اجرا افزایش خواهد یافت.آنها توان کمی مصرف می کنند ما فرض می کنیم که زمان نمونه برداری به خوبی کنترل می شود اما تاثیر ایده آل نبودن کلاک نمونه برداری و ایجاد تاخیر را مشخص می کنیم. زمانیکه تعداد سطوح سیم بندی شده محدود باشد فضا مابین بلوک ها برای خطوط دیتا بکار می رود.که این ناحیه به فضای حافظه اختصاص می یابد. بویژه زمانیکه سایز بلوک کوچک باشد. با توجه به اینکه تعداد سطوح سیم بندی در حال افزایش می باشد لذا در این مقاله فرض بر این است که در صورت زیاد بودن خطوط، خطوط دیتا به صورت عمودی نیز می توانند انتقال یابند. بنابراین فضای خطوط اضافی برای گذرگاه های افقی تنها در انتهای صف مورد نیاز می باشد.

تغییر ابعاد اندازه ترانزیستور روش دیگری برای سبک و سنگین کردن تاخیر، سطح مقطع و توان می باشد. در این مقاله فرض بر این است که اندازه گیت در مسیر انتقال به گونه ای است که باعث ایجاد کمترین تاخیر می شود. بنابراین fanout هرکدام از گیتهای منطقی به گونه ای اتخاب شده است که با تاخیردر گنجایش خروجی(fanout) چهار اینورتر بار شده برابر باشد. البته با این فرض،تاخیر کاهش نمی یابد ولی باعث کاهش

انرژی و کاهش سطح مقطع به میزان مطلوب می شود.یک مدلRC ساده که برای گیتهای منطقی بکار می رود. بنابراین اندازه گیت به گونه ای می باشد که کاهش تاخیر و یا افزایش تاخیر یکسان باشد پارامتر تک سایز wکه سایز ترانزیستور nMOS خواهد بود. اگر Cg ، خازن ورودی برای پهنای واحد و Rg مقاومت ورودی برای پهنای واحد اینورتر باشد مقاومت خروجی گیت برابر Ro=Rg/w و خازن ورودی برابر Cg_X $3w_X lc$ می باشد. مقدار تاخیر گیت منطقیw،بار C_Lرا از طریق مقاومتRw و ظرفیت خازنی Cw درایو می کند.(شکل 4)که در رابطه 1 بر آورد شده است.

$$D = \underbrace{R_0.(C_L + C_\omega)}_{\text{gate delay}} + p + \underbrace{R_\omega.\left(\frac{C_\omega}{2} + C_L\right)}_{\text{wire delay}} \qquad (1)$$

تاخیر گیت به علت ظرفیت خازنی نقطه اتصال درین می باشد.در طراحی SRAMپر بازده ،توان دینامیکی بر اتلاف توان کل چیره شده و بنابراین انرژی دینامیکی مورد نیاز برای سویچ ظرفیت خازنی در دکودر و mux خروجی را مدل می کنیم

سپس ما به بررسی جزئیات در دکودر وmux خروجی می پردازیم.

دکودر:

دکوراز دو جز تشکیل شده است :

دکودر سطر که خطوط کلمه ها را فعال می کند و دکودر ستون که سویچ ها را در خطوط بیت و mux خطوط دیتا مرتب می کند. از آنجائیکه دکودر سطر در مسیر دسترسی RAM قرار می گیرد لذا ما تاخیر آنرا مدل می کنیم در حینی که انرژی دکودر سطر و ستون را مدل می کنیم مسیر بحرانی دکودر با رشته ای از سه گیت منطقی که هرکدام شامل گیت NAND و اینورتر می باشد .مسیر دکود با اینورتر در ورودی با مینیمم سایز 16 λبرای pMOS و 8 λ برایpMOS درایو شده است.

تاخیر هر مرحله از مسیر دکود توسط فرمولی که در رابطه 2 نشان داده شده است محاسبه می شود. هر مرحله تاخیر برای حداکثر گنجایش خروجی 4 اینورتر به منظورکاهش تاخیراندازه گیری شده است زمانی که مقاومت سیم predecode و خطوط کلمه قابل صرفنظر نباشد (در شکل 5) بافرهای اضافی برای کاهش تاثیر بارهای گیت ورودی مورد نیاز خواهد بود. تعداد مطلوب بافر با محاسبه تاخیر دکودبا تعداد بافرهای مختلف در این دوقسمت قابل محاسبه می باشد.

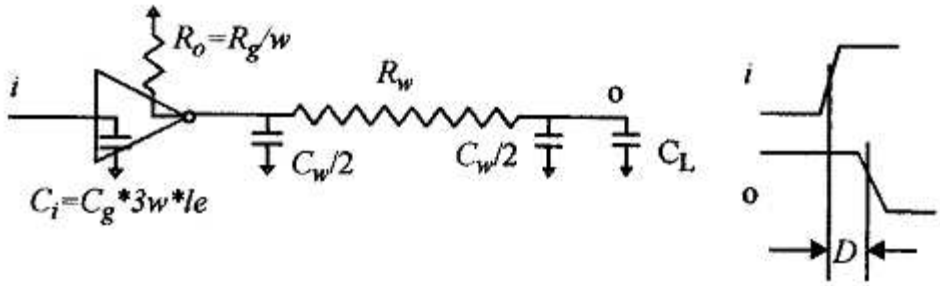


Fig. 4. Delay of a logic gate driving a load through an RC line.

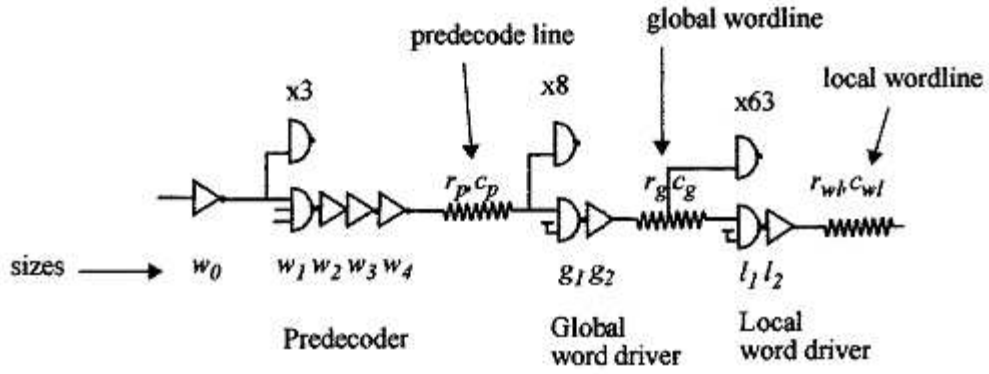


Fig. 5. Model of the critical path of a 12–4096 row decoder.

تاخیر دکودر با حداکثر گنجایش4 سایز در قسمت 2 بیان شده است .

$$D_{decoder} = N.\tau fo Ax + \sum_{i=0}^{N} pi + D_{wire} \qquad (2)$$

درایورهای کلمه به سیم های predecode متصل شده است. وجود بار به علت تمامی ورودی هایی که به خط کلمه سراسری وصل شده است ، می باشد. در یک SRAM واقعی، درایورهای کلمه محلی به صورت یکنواخت توزیع شده است .

برای کاهش تاخیر سیم، خط کلمه سراسری را که از مرکز سیم درایو شده است که دو بخش به صورت موازی درایو شده بر

روی آن تاثیر می گذارد که هر بخش دارای مقاومت rg/2 و ظرفیت خارنی برابر (Cg+l)/2 می باشد.تاخیر در سیم خط کلمه سراسری کلی برابر 0.5x(Cg+l)/2x rg/2 می باشد.

تاخیر سیم خالص برابر با:

$$D_{wire} = r_p.(C_p/2 + G) + r_g(c_g + L)/8 + r_{wl}.C_{wl}/8 \qquad (3)$$

تاخیر برای دکودر سطر در چهار SRAM مختلف 9٪ تاخیر شبیه سازی درHSPICE می باشد.(شکل6)

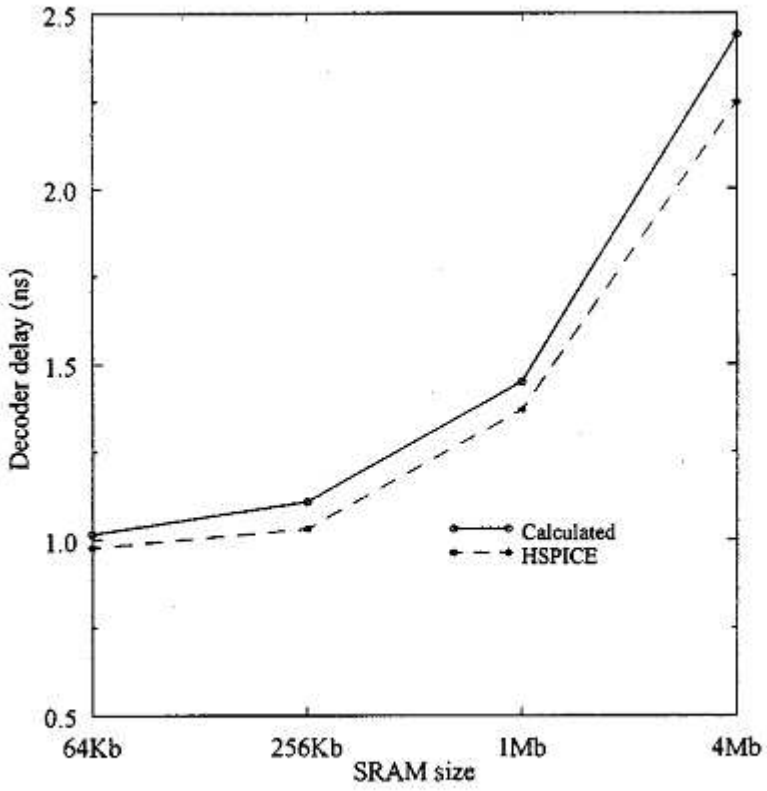


Fig. 6. Comparison of estimated and HSPICE simulated delay for row decoders.

تاخیر خط بیتی به زمان خیز خط کلمه محلی بستگی دارد،لذا edge rateدر انتهای خط کلمه را محاسبه می کنیم. از شبیه سازی مداری زمان خیز برابر با 1.4 برابر تاخیر مرحله آخر درایور کلمه بدست امده که به صورت خلاصه در 4 قابل ملاحظه است:

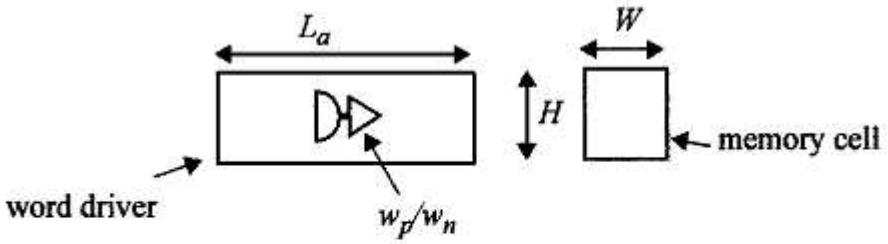
$$\text{rise time} = (\tau_{fo4} + r_{wl}.C_{wl}/8) \times 1.4 \qquad (4)$$

ظرفیت خارنی گیت و سیم در مسیر سیگنال به منظور محاسبه انرژی دکودر اضافه شده است.سطح دکودر با استفاده از

محاسبه سطح درایورهای کلمه محلی و سراسری و سطح مورد نیاز برای سیم های predecode محاسبه می شود.سطح

درایورهای کلمه به عنوان تابع خطی از پهنای کلی دستگاه داخل درایور مدل می شود. (شکل7) مقدار ثابت برای این تابع

از آرایش شش درایور کلمه مختلف به دست می آید که برابر λ^2 نمی باشد که λ نصف مینیمم سایز طرح تکنولوژی می باشد.



$$L_a{*}H = 24.05{*}1.25 * (w_p + w_n) + 497 \quad (in\ \lambda^2)$$

Fig. 7. Area estimation for the word drivers. The constants have been obtained by an empirical fit on areas from actual layouts.

پهنای کلی دستگاه 1.25 برابر بافر نهائی می باشد. ناحیه کاری predecode عمودی در بلوک سیمی (شکل1) همواره به ناحیه دکود کلی اضافه می شود. به عنوان مثالی افزایش درپهنای آرایه (صف) SRAM با استفاده ازسطح برای درایور کلمه محلی 64، یک درایور کلمه سراسری64 و سیم بندی عمودی برای 16سیم predecode و 64 بلوک سیمی انتخابی محاسبه می شود.

MUX خروجی

Mux خروجی شامل ماکس خط بیتی می باشد که سلول داده رادر تقویت کننده قرار می دهد و ماکس خطی دیتا ،دیتا را از تقویت کننده حسی به خروجی انتقال می دهد. از آنجائیکه سطح سیگنال در هر دوی این ماکس ها کوچک می باشد. (~100 mV) منبع سیگنال ورودی برای هر دو ماکس منبع جریان ایده آل مورد استفاده قرار گیرد.کاهش تاخیر از طریق شبکه RC برای منبع جریان ورودی متفاوت از منبع ولتاژ ورودی می باشد منبع جریان ایده آل که برای درایو شبکه RC بکار می رود (شکل .(8(a) شکل موج ولتاژ در گره 1و3 در شکل 8(b) نشان داده شده است. ثابت زمانی شبکه در رابطه 5 نشان داده شده است .

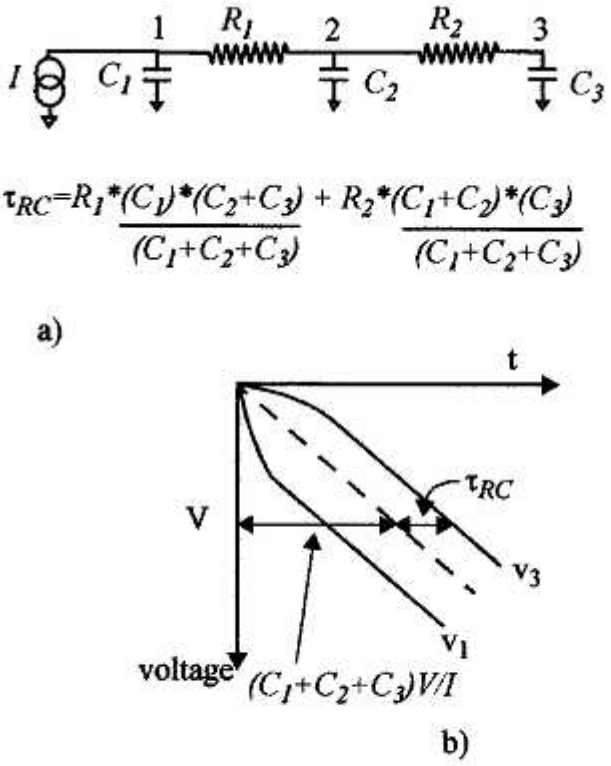


Fig. 8. (a) Current source driving a RC π network and (b) sketch of the node waveforms.

در حالت ثابت ($t>>\tau_{RC}$) گره 1و2و3 slew rate یکسانی دارند و تاخیر برای نوسان ولتاژ در گره3 با فرمول 6 به صورت تقریبی قابل محاسبه ست که تاخیر را برای زمانی می باشد که مقاومتی جود ندارد .این فرمول برای محاسبه هر دو ماکس خط بیت و خط دیتا بکار می رود

$$D + \frac{(C1+C2+C3).V}{I} + \tau_{RC} \qquad (6)$$

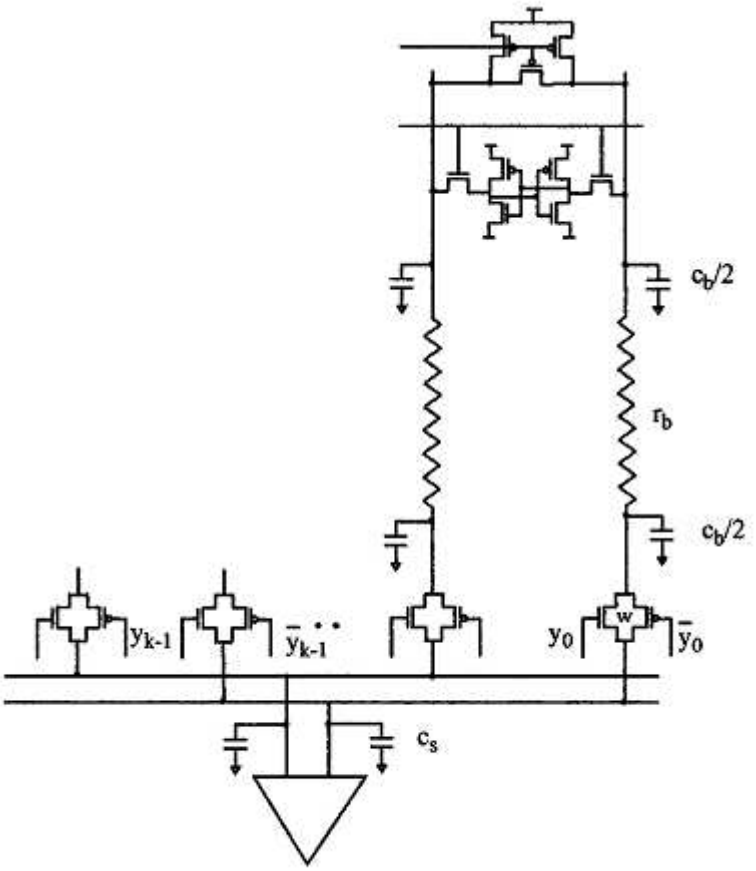


Fig. 9. Schematic of a single-level bit line structure.

یک ماکس خط بیتی تک سطحی در شکل 9 نشان داده شده است و به عنوان یک منبع جریان ایده آل برای درایو شبکه
RC (شکل8) مدل شده است . سویچ ماکس و سیم خط بیتی محلی و سراسری باعث افزایش مقاومت و ظرفیت خازنی شبکه می شود. تاخیر خط بیتی برای نوسان سیگنال توسط σ_v مجموع تاخیر برای ایجاد نوسان ولتاژ با ثابت زمانی شبکه RC و بدون در نظر گرفتن مقاومت می باشد. خط کلمه محلی طولانی به علت مقاومت خط دارای زمان خیز کندی می باشد از آنجایکه زمان خیز تاخیر سلولی را تحت تاثیر قرار می دهد لذا باید در مدل آنرا در نظر بگیریم. تاثیر زمان خیز (Tr) با اضافه کر دن عبارت اضافی به معادله تاخیر که متناسب با آن می باشد،بدست می آید. ثابت تناسبی به نسبت ولتاژ threshold سلول به ولتاژ منبع بستگی دارد و در شبیه سازی مقدار 0.3را برای رنج وسیع بدست آورده ایم. ثابت زمانی در معادله تاخیرخط بیتی در عبارت 5 محاسبه شده است.

$$D_b = (C_b + j_{mux}.(K+1) + Cs).\frac{\delta v}{I_{Cell}} + z.T_r + \tau_{RC} \quad (7)$$

- ظرفیت خارنی خط بیتی C_b
- ظرفیت خارنی اتصال قسمت سویچ ماکس j_{mux}
- تعداد ستون مالتی پلکس شده یه تقویت کننده حسی منفرد K
- ظرفیت خارنی ورودی تقویت کننده حسی Cs
- ولتاژ نوسان در ورودی تقویت کننده حسی σ_v
- جریان سلول حافظه Icell
- زمان خیز خط کلمه محلی Tr
- ثابت تناسبی مشخص شده بوسیله HSPICE z
- ثابت زمانی خط بیتی در شبکه RC τ_{RC}

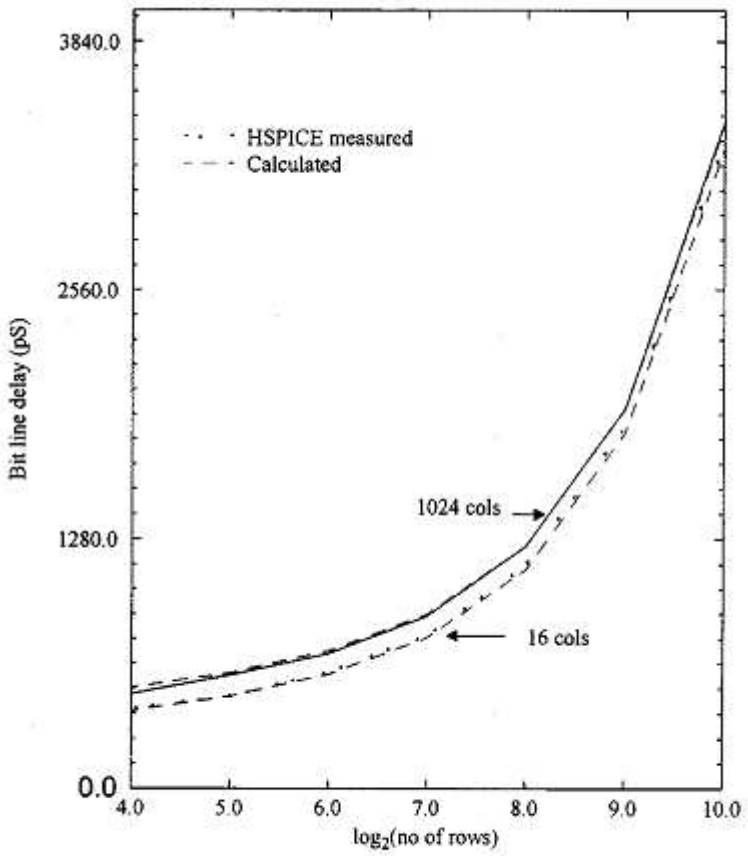


Fig. 10. Bit line delay versus column height; 0.25 μm, 1.8 V, and four columns multiplexing.

شکل 10 نشان دهنده ی تاخیر اندازه گیری شده توسط HSPICE از طریق درایور خط محلی و خط کلمه مقاومتی و خط بیتی ، نسبت به ورودی تقویت کننده حسی می باشد. تاخیر حساب شده برابر 2.4٪ تاخیر HSPICE زمانیکه ارتفاع خط بیتی 32 ردیف از هر دو ردیف کوتاه (16ستون) و بلند(1024ستون) می باشد .

بافر تقویت کننده حسی که در شکل 11 نشان داده شده است شامل لچ های cross-coupled اصلی که توسط زنجیره ای از اینورترها و یک جفت از درایورهای nMOS به هم وصل شده اند. مبدل لچ سیگنال نوسانی ورودی کوچک را به سیگنالCMOS نوسانی کامل تبدیل کرده و از آن برای هر دو تقویت کننده حسی محلی و سراسری استفاده می کند. در تقویت کننده حسی محلی ،خروجی لچ توسط زنجیره ای از اینورترها و به گیت خروجیnMOS درایو شده است. این ترانزیستورهای nMOSیک سیگنال ولتاژ کوچکی در خروجی آنها با دشارژکردن خط دیتا بوجود می آورند. تاخیر در ساختار تقویت کننده حسی برابر با مجموع تاخیر تقویت کننده لچ τ_S و تاخیر بافر و درایور خروجی می باشد.τ_Sمتناسب با لگاریتم گین ولتاژهای مطلوب و بارخروجی تقویت کننده می باشد. .اگر فرض کنیم تمام ترانزیستورهای لچ در همان مقیاس تناسبی باشند، مقاومت خروجی و

ظرفیت خازنی ورودی با تابعی از سایز nMOS که cross-coupled در لچ می باشد، بیان میشود. (شکل 11) درایورهای nMOS به عنوان منبع جریان مدل می شوند که خروجی جریان آنها متناسب با اندازه ی ωn می باشد.در دکودر سایزهای $w_s w_1 \ldots w_n$ برای کاهش تاخیر ماکس خروجی کلی بکار می رود. معادله 8 نشان دهنده ی مقدار مناسب تاخیر ماکس خروجی برای بهینه سازی می باشد و برابر با مجموع تاخیرهای ماکس خط بیتی،تقویت کننده حسی لچ،بافرها،درایورهای nMOS می باشد.

$$T = D_b(w_s) + \tau_s + \frac{R_s \cdot 3 \cdot C_g \cdot w_1}{w_s} + \frac{R_g \cdot 3 \cdot C_g \cdot w_2}{w_1} \cdots \frac{R_g \cdot C_g \cdot w_n}{w_{n-1}} + \frac{C \cdot \delta v}{I_n \cdot w_n} + \text{other constants} \qquad (8)$$

$$D_b(ws) \approx \frac{Gs.Ws.\delta v}{Icell} \qquad (9)$$

- $\tau s = 2\tau f_{04}$ o
- Gs=3.8fF/λ o
- Rs=36k o
- w_s سایز تقویت کننده حسی o
- Cg,Rg, مقاومت خروجی و ظرفیت خارنی ورودی o
- In=3.75μA/λ o
- C ظرفیت خارنی ماکس خط دیتا o

برای مشخص کردن روشی برای یافتن مطلوبترین سایز، تاثیر سایز تقویت کننده حسی لچ بر روی ثابت زمانی ماکس خط بیتی نادیده گرفته شده و تنها تاخیر سلولی مورد بررسی قرار گرفته (9) به عبارت ساده تراز تاثیر ظرفیت خارنی اتصال nMOS بر روی ثابت زمانی RC صرفنظر شده است.هر دوی این عاملها تاثیر کمی بر روی بهینه سازی سایز دارند ولی برای محاسبه نتیجه نهایی انها را در نظر می گیریم.مینیمم تاخیر از طریق ساختار تقویت کننده حسی زمانی صورت می گیرد که هر کدام از عبارتهای در (8) برابر با تاخیر خارجی در4 اینورتربارشده باشد.تاخیر در تقویت کننده حسی در طرحی مشابه محاسبه می شود با این تفاوت کهدر این تحلیل، تاخیر بافر به درایو بار خروجی بررسی نشده است .

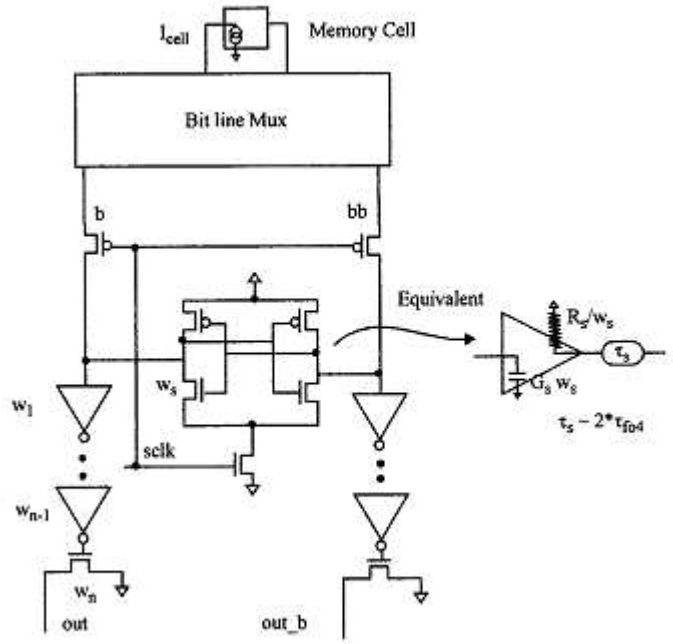


Fig. 11. Local sense amplifier structure.

برای عملیات کم توان،سیگنال در گره ظرفیت خارنی بالا،با نوسان ولتاژ کمی نگه داشته می شود.خط بیت با فرستادن خط کلمه باعث نوسان سیگنال کلی برابر $2\sigma_v$ می شود.بنابراین انرژی ماکس خط بیت و ماکس خط دیتا که شامل ظرفیت خارنی خط و سیم و اتصالات و ظرفیت خارنی گیت ورودی و ولتاژ منبع می باشد محاسبه می شود.انرژی هر قسمت از شبیه سازی برابر 12 fJبرای 0.25- m پروسه می باشد. سطح سویچ در ماکس خط بیتی وتقویت کننده حسی مداری و precharge و درایورهای نوشتاری به قسمت عمودی ردیف SRAM اضافه می شود.(شکل12) از آنجائیکه درایور نوشتن، precharge و ترانزیستور ماکس بهینه سازی نشده اند بنابراین به ترتیب یک سرجمع ثابت 4، 1و2 سلول حافظه اضافه می کنیم. سطح تقویت کننده حسی محلی به عنوان تابع خطی از پهنای دستگاه کلی داخل تقویت کننده حسی می باشدپارامترها برای مدل از 5 طرح مختلف بدست آمده است و در شکل 12نشان داده شده است. پهنای دستگاه کلی داخل تقویت کننده حسی با پارامترهای ω_S،ωn و ω_Pمحاسبه می شود. مجموع پهنای دستگاه داخل لچ برابر $8.7_{\times}w_S$ که مقدار8.7 از طرح لچ بدست آمده است. با گنجایش خروجی 4 سایز، سطح فعال بافر بیشتر از 1/3پهنای درایور نمی باشد. از این رو سطح فعال 2 درایور nMOS و بافرهای آنها برابر با$1.33\times2\times\omega n$می باشد. سپس ما به بررسی نتایجی که با استفاده از این مدلها برای آنالیز ساختار RAM متعدد در سایزهای مختلف و تکنولوژی ساخت متفاوت می پردازیم.

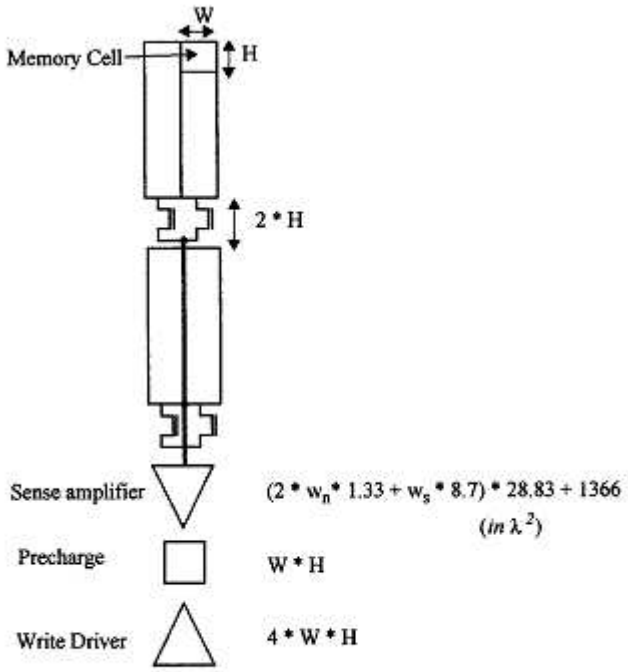


Fig. 12. Area estimation of the output mux.

نتایج آنالیز:

ما تمامی ساختار هایRAM را یکا یک شمرده و سطح وتاخیر و انرژی هر کدام را با استفاده از مدل ساده ای در قسمت قبلی توضیح دادیم.که مشخص کردن ساختارهای بهینه برای توابع تاخیر،سطح و انرژی را و کمینه ساختن مقادیر آنها را فعال می کند.

$$\min (1 - \beta - \alpha) \cdot \text{Delay} + \alpha \cdot \text{Area} + \beta \cdot \text{Energy}. \qquad (10)$$

منحنی تغییرات با عوض کردن مقادیر وزنی α وβ که بین 0 و 1 می باشد بدست می آید.شکل 13 نشان دهنده ی تاخیر ساختار SRAM برای حداقل تاخیر (0=β= α) می باشد که به دو صورت با در نظرگرفتن مقاومت سیم و بدون آن ، برای سایز 64 kb تا 16 Mb با پهنای دسترسی 64بیت در تکنولوژی 0.25-um نشان داده شده است. تاخیر SRAM بدون مقاومت سیم برابر 15 τf_{04} برای طرح 64-kb و متناسب با لگاریتم ظرفیت مشاهده شده در می باشد. تاخیر برای هردو RAM تا $1.2\tau f_{04}$ افزایش می یابد .تاخیر برای هر دوی اینها در همان نمودار نشان داده شده و برابر با مقدار بهینه سازی شده برای ساختار SRAM می باشد.

برای دکودر سطر هر بیت آدرس نصف ردیف را انتخاب می کند و بار دیده شده توسط بیت آدرس متناسب با S/2می باشد.که S تعداد کلی بیتهای ردیف می باشد. با دوبرابر کردن در تعداد بیتها نصف تاخیر τf_{04} اضافه می شود.در مسیر خروجی ،با دوبرابر کردن سایز خازن سیم در ماکس خط دیتا 1.4برابر افزایش می یابد.که به محیط آرایه متناسب می باشد.بنابراین تاخیر تقویت کننده حسی محلی در حدود $0.25\tau\ f_{04}$ افزایش می یابد.منحنی نهائی در شکل 13 تاخیر با مقاومت سیم را نشان می دهد که سیم های سراسری برای این منحنی پهنای/mm 7.5دارا می باشند. تاخیر RC سیم با طول سیم افزایش می یابد .تاخیر سیم برای سیم های سراسری در مقیاس SRAM برابر با سایز SRAM می باشد و برای SRAM بزرگ بیشتر نمایان می باشد.

بهینه سازی پهنای سیم به منظور کاهش تاثیر تاخیر داخلی صورت می گیرد. شکل 14تاخیر کلی را برای-4 SRAM Mbدر دو پهنای سیم مختلف در چهار تکنولوژی ساخت مختلف نشان می دهد. فرض بر این است که مس با ضخامتm 0.18می باشد .منحنی پایینی نشان دهنده ی تاخیر با مقاومت صفر برای سیم می باشد. همانطوری که در شکل مشاهده می شود برای 0.1 µm تاخیر$2.2\tau\ f_{04}$ وبرای 0.07µm تاخیر $3.6\ \tau\ f_{04}$ افزایش می یابد زمانیکه از تاخیر داهلی صرفنظر شده است.منحنی دوم تاخیر سیگنال round-trip را در مسیر دسترسی با فرض سرعت نور متناسب با1 mm/6.6 pS و مقدار تاخیر داخلی کمتری را ایجاد می کند. سرعت

تاخیر داخلی نور برابر $1.75\tau f_{04}$ برای 4-Mb SRAM می باشد که مستقل از تکنولوژی می باشد وبرای هر جهار برابر سایز RAM ، دوبرابر افزایش می یابد . دو منحنی بالای آن نشاندهنده ی تاخیر با در نظر گرفتن مقاومت غیر صفر برای دو پهنای سیم 8λو10λ می باشد .

کاهش تاخیر سیم زمانی امکانپذیر است که سیم های کلفت برای سیمبندی سراسری بکار رود. از$0.25\ \mu m$ با سیم بندی آلومینیومی تا$0.18\ \mu m$ با سیم بندی مسی لزوما باعث تاخیر می شود.اما با طرح جمع شدن اضافی (further shrinks)،تاخیر برای تمامی پهنای سیمها کاهش می یابد.

تاخیر RC سیم به خوبی تاخیر گیتمقیاس بندی نشده است.به هر جهت با پهن سازی سیم ،امکان حفظ تاخیر قبلی میسر می شود.پهنای سیم10λ تاخیر با τf_{04} سرعت نور محدود بین $0.25\text{-}\ \mu m$ و 0.18 امکانپذیر می کند. تا زمانی که سیم های پهن تر $0.1\ \mu m$ تا $0.07\ \mu m$ مورد نیاز می باشد .

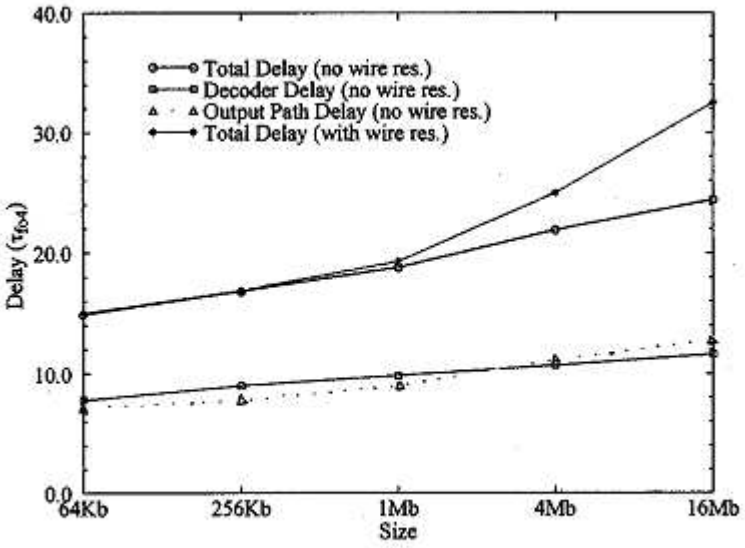


Fig. 13. Delay scaling with size in the 0.25-μm process.

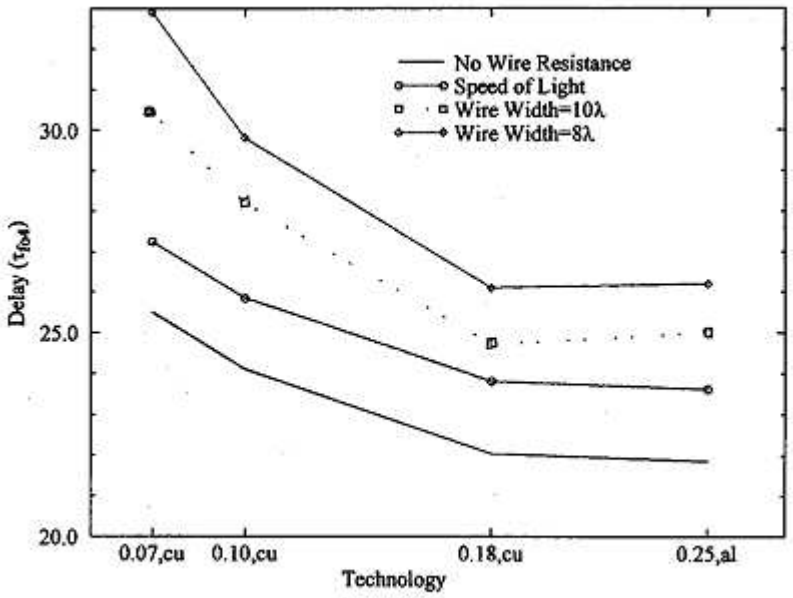


Fig. 14. Delay versus technology for different wire widths for a 4-Mb SRAM.

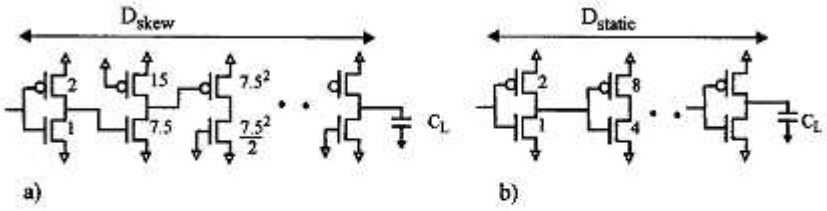
سپس ما روش هایی که عملکرد اجرا و پیاده سازی SRAM واقعی را که ممکن است با منحنی های قبلی متفاوت باشد بررسی می کنیم. معمولاSRAM بزرگ مدارهای تکراری در مسیر دکود را با هم ترکیب می کند که باعث ایجاد منحنی تاخیربرای شیفت از 1/2تا τf_{04} می شود. SRAM های پر بازده از سبک مداری استاتیک در مسیر دکودر استفاده نمی کنند. برای محاسبه افزایش سرعت با skewing، ابتدا زنجیره ای از اینورترها که سیگنال ورودی یا به گیتnMOS یا به pMOS متصل شده باشد را در نظر می گیریم. (شکل15) با فرض اینکه MOSFET مکمل در دسترس و گیت آن غیر فعال شده باشد. (در راه اندازی واقعی در ری ست کردن فاز ،فعال می شود) و آن فقط به گیت با بارخود پرشونده اضافه می شود. با در نظر گرفتن این فرض ها و پارامترهای جدول 1حداکثر گنجایش خروجی مطلوب برای در حدود 5و تاخیر گیت skewed تقریبا برابر 70٪ گیت بدون skewed می باشد. اگر predecoder و درایور کلمه سراسری بصورتskewed باشند تاخیر دکودربرای اجرای استاتیک به جای$8\tau f_{04}$ به $6\ \tau f_{04}$ کاهش می یابد. با دو برابر کردن سایز RAM، تاخیر دکودر تا$0.3\ \tau f_{04}$به جای $0.5\ \tau f_{04}$افزایش می یابد.سرانجام کلاک برای تقویت کننده حسی معمولا با تاخیر برای فعال کردن عملکرد در رنج وسیعی صورت می گیرد که موجب تاخیر اضافی تا τf_{04} شود. قسمت بندی کردن برای سبک و سنگین کردن میان تاخیرها ،سطح و انرژی بکار می رود.منحنی تغییرات نشان دهنده ی مقادیر مختلف برای $\alpha\beta$ می باشد که اگر β صفر شود منحنی برای SRAM 4-Mb در پروسه 0.25- mدر شکل 16 نشان داده شده است. هر نقطه بر روی این منحنی نشان دهنده ی کمترین سطح قابل

دسترسی از طریق ساختار جدید RAM برای تاخیر مشابه می باشد. با شروع از کمترین تاخیر که به صورت دقیق قسمت بندی شده است، بهبودی مطلوب در سطح با کاهش مقادیر قسمت بندی و قبول تاخیر کوچک امکانپذیر می باشد و کاهش متوالی در قسمت بندی باعث کاهش بهبودی در سطح برای پذیرش تاخیر می شود. پارامترهای قسمت بندی در نقاط A، B و C در شکل نشان داده شده است. نقاط B،A قسمت مرکزی منحنی را تشکیل می دهند که در مقایسه با اجرای سریع، A ،22٪ کندتر و 22٪سطح کوچکتر و B 14٪ کندتر و 20٪ سطح کوچکتر دارند.

اغلب پارامترهای ساختاری مختلف ،تاخیر RAM بسیار حساس به ارتفاع بلوک می باشدو زمان دسترسی سریع با استفاده از ارتفاع کم بلوک میسر می شود.شکل17 تاخیر و سطح را برای 4-Mb SRAM با ارتفاع بلوک های مختلف نشان می دهد زمانیکه از مقادیر مطلوب برای پارامترها استفاده شده است.ارتفاع کم بلوک تاخیر خط بیتی را کاهش ولی تاخیر سیم سراسری را افزایش می دهد. برای ارتفاع بلوک بلند تاخیر خط بیتی محدود به زمان دسترسی می شود.بنابراین ارتفاع مطلوب برای این مثال برابر 32ردیف می باشد.افزایش ارتفاع به 128 ردیف منجر به افزایش تاخیر در حدود 8٪ در صورتی که سطح تا 7.6٪ می تواند کاهش یابد.

با قرار دادن $\alpha = 0$در معادله10 میزان انرژی تاخیر از طریق قسمت بندی ،بدون در نظر گرفتن محدودیت سطح بدست می آید که در شکل 18 نشان داده شده است.قسمت عمودی سمت چپ میزان انرژی مصرفی برای سویچ گیت اینورتر می باشد (Eunit 72 fJ) قسمت بندی امکان مقایسه بین انرژی و تاخیر را فراهم می کند . شکل میزان مطلوب مالتی پلکس ستون (cm) و ارتفاع بلوک (bh) برای پاسخ تاخیر و انرژی برای تعدادی نقاط نشان می دهد.

به این نتیجه رسیدیم که راه حل برای انرژیهای کم استفاده از یک ستون برای مالتی پلکس می باشد.از آنجائیکه ما بهینه سازی برای کاهش تاخیر انجام می دهیم ترانزیستورهای آخری در خروجی تقویت کننده بزرگ می باشند و دارای ظرفیت خارنیبزرگی می باشد.از این رو در طراحی کم انرژی داشتن ارتفاع زیاد بلوک مزیت به شمار می رود که باعث می شود اکثریت muxing در ماکس خط بیتی که ظرفیت خارنیs ترانزیستور سلول دسترسی حافظه بسیار کمتر از ظرفیت خارنیs در ماکس خط دیتا می باشد ،صورت گیردهمچنین مصرف انرژی در بهینه سازی ساختار SRAM به مجموع دو جز بیان می شود که یکی مستقل از ظرفیت می باشد تنها به پهنای دسترسی بستگی دارد وبه علت خط کلمه محلی،سیگنال precharge ،تقویت کننده محلی و سراسری و... مقیاس قسمت دیگری ریشه دوم ظرفیت می باشد همانطوری که در . مشاهده شدبا اتلاف قدرت در سیم های سراسری و دکودر ها ارتباط دارد.



$$D_{skew} = 0.31 * \tau_{fo4} + 0.62 * D_{static} \qquad (10)$$

Fig. 15. Optimal sizing for (a) extremely skewed and (b) statically sized inverters.

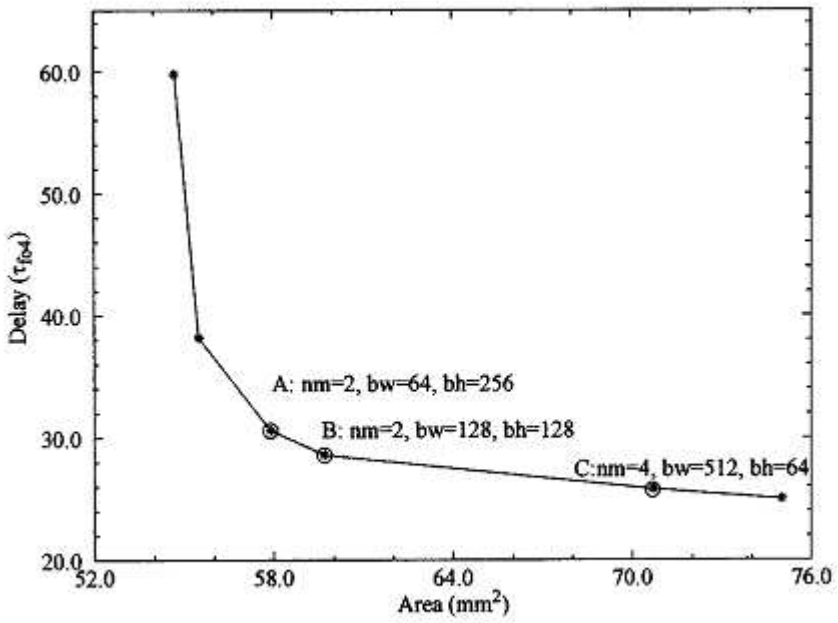


Fig. 16. Delay versus area for a 4-Mb SRAM in the 0.25-μm process.

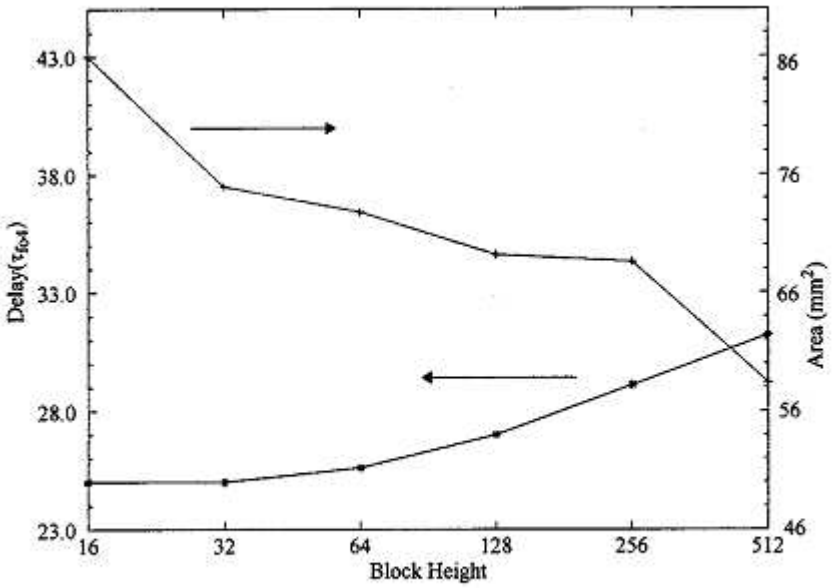


Fig. 17. Delay and area versus block height for a 4-Mb SRAM in a 0.25-μm process.

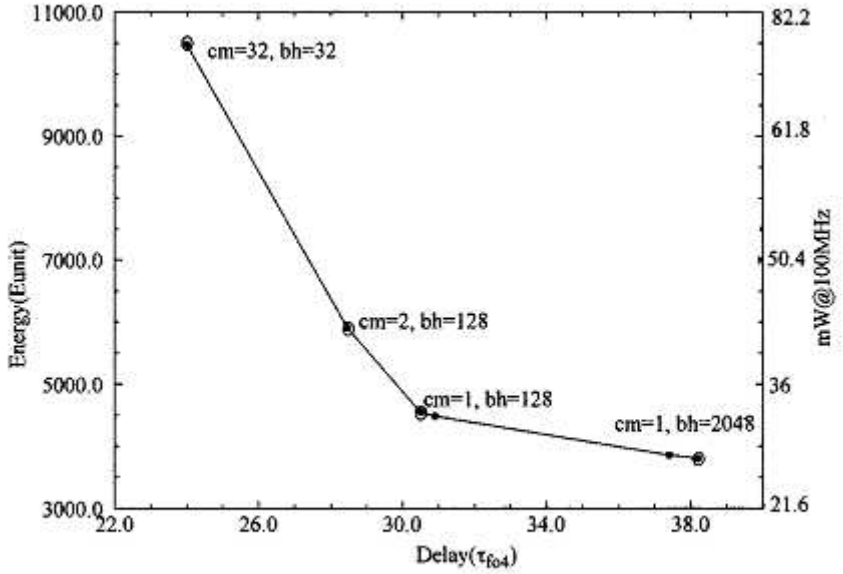


Fig. 18. Energy versus delay for a 4-Mb SRAM in a 0.25-μm process.

نتیجه گیری :

مدل های تحلیلی برای تاخیر، سطح و انرژی ، طراحی در محدوده کوچک زمانی را امکانپذیر می سازد. این مدل ها برای بررسی تاثیر قسمت بندی SRAM بکار می رود. تغییرات قابل توجه بین سطح،تاخیر و انرژی از طریق انتخاب ساختارهای SRAM قابل مشاهده است. این مدل ها برای پیش بینی میزان تاخیر با ظرفیت و تکنولوژی پروسه بکار می رود. تاخیر SRAM به دو قسمت قابل تفکیک می باشد.یکی به علت ترانزیستورها ی بکار رفته در تکنولوژی می باشد. (تاخیر گیت) و دیگری به خاطر اتصال درونی (تاخیر سیم) می باشد. تاخیر گیت به ازای دو برابر کردن سایز RAM، $1.2\tau f_{04}$ افزایش می یابد.زمانیکه از طرح مداری استاتیکبرای طراحی دکودر استفاده می شود برای 64-kb RAM از $15\tau f_{04}$ شروع می شود. برای بهینه سازی ساختار 4-Mb SRAM، افزایش تاخیر در حدود $2\tau f_{04}$ برای 0.1-μ m و در حدود$3.6\tau\ f_{04}$ برای0.07-μm می باشد که در ساختار دیگر کمتر می باشد. برای اکثر ساختارهای SRAM با استفاده از طرح های مرتبه ای برای مسیرهای خط بیتی و خط دیتاو با استفاده از تکنیکهای جبران آفست،می توان از افزایش تاخیر جلوگیری کرد.تاخیر سیم برای RAMهای بیشتر از 1-Mb تاثر گذار می باشد.با استفاده از روش shrink،تاخیر سیم کاهش می یابد و طراحی دوباره سیم برای حفظ تاخیر سیم نسبت به تاخیر گیت با همان نسبت قبلی باید انجام گیرد.تقسیم ساختار خط کلمه برای دکودرها و ماکس ستون برای مسیر خط بیتی ،فضای کافی برای استفاده از سیمهای کلفت تر ایجاد می کند که باعث کنترل تاخیر برای4-Mbو طراحی کوچکتر می باشد.تاخیر سیم دارای حد کمتری بوسیله سرعت نور می باشد که برای SRAM 4 –b در حدود $1.75\tau\ f_{04}$ می باشد و با جهار برابر کردن ظرفیت ،دو برابر می شود.در سطح های بالا تر، معماری RAM نیاز به تغییر برای استفاده از مسیر یابی آدرس و دیتا دارد. (مثال14 را مشاهده کنید.) تاخیر سیم متناسب با سطح سلول می باشد و طراحی سلول با سطح کمتر برای RAM بزرگ مزیت به شمار می رود.حتی اگر سلول ضعیفتر باشد. بنا براین سلول DRAM،سلول چند ارز multivalued ، سلول TFT-based و طراحی سلول های جدید دیگر برای کارایی زیاد و ظرفیت زیاد RAM مورد نیاز می باشد.

ضمیمه :

ما تمام فرض هایی را که در قسمت مدل سازی بیان نشد را در این قسمت لیست کرده ایم :

تکنولوژی :

تکنولوژی اصلی در مقیاس0.25µm CMOS فرض شده است و توضیحات مناسب در جدول 1 نشان داده شده است. ویژگیهای اساسی برای 4ساختار مختلف در جدول 2 بیان شده است.فلز کاری مس با فرض تولید 0.18-µ m به بعد انجام شده است.فلز سطح بالاتر به حالت ضخیم تر طراحی شده است: ارتفاع و پهنای آنها در مقیاسی طراحی شده است که دارای قسمت عبوری بزکتری باشند اما افزایش ارتفاع آن تنها با ریشه دوم فاکتور افزایش پهنا امکانپذیر می باشد. برای مثال لایه فلزی سطح بالا تر با دو برابر پهنای مینیمم لایه فلز 1 دارای ارتفاع 1.4برابر ارتفاع فلز 1 می باشد بنابراین مقاومت فاکتور 3 کوچکتر از مقاومت فلز 1 می باشد.فرض ما بر این است کهpitch سیمبندی دوبرابر پهنای سیم برای سیم های سراسری می باشد.

معماری :

SRAM سنکرون می باشد برای مثال کلاک امکان دسترسی را شروع می کند بنابراین به آسانی نتیجه می شود که با اضافه کردنSRAM آسنکرون توان و تاخیر برای ردیابی انتقال آدرس (ATD) بکار می رود.ساختار SRAM جا سازی شده مفروض است. تمام بیتهای دیتا کلمه ی دسترسی از هسته حافظه در مجاورت فیزیکی یکدیگر حاصل می شود (شکل1) برخلاف SRAM مستقل، محل قرارگیری پورت IO دیتا برای ارتباط خارجی بهینه سازی شده است.این بهینه سازی یک مقدار ثابت آفست به تاخیر و توان هسته SRAM اضافه می کند که نتیجه این مطالعه قابل اجرا برایSRAM مستقل می باشد . اندازه سلول برای تحلیل در نشان داده شده و سطح سلول معمولا از طرح بندی 6ترانزیستور CMOS با چگالی بالا می باشد.

ساختار مدار :

RAM برای کارکرد پرسرعت با توان کم طراحی شده است که اتلاف انرژی را بدون تاثر بر سرعت کاهش می دهد. خط کلمه محلی برای کنترل نوسان خط بیتی بکار می رود که نوسانهای کم برای کاهش توان در خط دیتا بکار می رود.از آنجایی که این روش سرعت RAM را تحت تاثیر قرار نمی دهد نتایج تحلیل مامربوط به میزان تاخیر قابل اجرا برای طراحیSRAM با سرعت مطلوب می باشد.یک تقویت کننده حسی طرح لچ با کنترل زمانی کامل برای تقویت کننده فرض شده است که توان بسیار کمی مصرف می کند و کارایی سریع دارد.بنابراین نتایج تحلیل ما منجر بهSRAM پرسرعت و کم توان می شود.برای پروسه 0.25- µm نوسان

ورودی مطلوب که تاخیر تقویت کننده را کاهش می دهد از شبیه سازی مقدار100 mV که50 mV آفست ورودی می باشد.ترانزیستورها در ماکس خط بیتی سایز ثابتی دارند و در ماکس خط دیتا برای مشخص کردن تحلیل پهن تر می باشند .شبیه سازی مداری تاخیرRAM را نشان می دهد که حساسیت کمی به سایز این ترانزیستورها دارند.

مدل انرژی :

نوسان در خط بیتی و خط IO به عملیات کم توان محدود می شود.زمانی که دقیقا به مقدار مطلوب توسط تقویت کننده حسی محدود می شود،در طراحی عملی،متغیرهای کمکی برای کنترل مناسب آن وجود دارد و فرض می کنیم که دوبرابر نوسان سیگنال مطلوب باشد.بنابراین برای پروسه 0.25- µm،نوسان برابر می باشد که200 mV می باشد که نوسان مطلوب برای تقویت کننده برابر .100 mVمی باشد.