Prevod iz časopisa IEEE Journal of Solid State Circuits, Vol. 27, No. 4, April 1992, pp 475-483
LOW POWER CMOS DIGITAL DESIGN

Chandrakasan A., Sheng S., Brodersen R.

-Projektovanje kola sa malom potrošnjom-

I. UVOD

Poslednjih godina uložen je veliki napor usmeren ka istraživanju metoda za povećanje brzine digitalnih sistema, tako da zahvaljujući današnjim tehnologijama postoje veoma moćne personalne radne stanice, koje se odlikuju sofisticiranom kompjuterskom grafikom i multimedijalnim sposobnostima kao što su prepoznavanje govora u realnom vremenu i video u realnom vremenu. Veoma brzo izračunavanje postalo je tako, standardno za prosečnog korisnika a ne samo privilegija pojedinaca koji imaju pristup moćnim *mainframe* računarima. Pored toga, druga značajna promena u stavu korisnika jeste želja da se ostvaruje pristup izračunavanju sa bilo koje lokacije bez potrebe povezivanja na kablovsku mrežu. Zahtevi za prenosivošću tako postavljaju nekoliko ograničenja koja se pre svega odnose na veličinu, težinu, i potrošnju energije. Potrošnja je posebno bitna obzirom da standardne nikl-kadmijumske baterije obezbeđuju samo 40 Wh energije po svakom kilogramu težine. Iako su prisutna stalna poboljšanja u tehnologiji proizvodnje baterija, neosporno je da se nameće veliki problem vezan za potrošnju energije.

U početku su se prenosive aplikacije odlikovale malom potrošnjom i malom propusnošću, kao npr. digitalni ručni satovi i kalkulatori, međutim danas je sve veći broj aplikacija koje zahtevaju malu potrošnju i veliku propusnost. Npr. *notebook* i *laptop* kompjuteri predstavljaju segment industrije koji se najbrže razvija, i koji zahteva sposobnost izračunavanja istovetnu kao kod *desktop* mašina. Podjednako je zahtevan i razvoj personalnih komunikacionih sistema (PCS), kao što je to slučaj sa mrežama digitalne mobilne telefonije koje se zasnivaju na kompleksnim algoritmima za kompresiju govora i sofisticiranim radio modemima minijaturnih dimenzija. Takođe, značajna su i nastojanja da se bežičnim linkovima prenose i drugi tipovi informacija kao što su *full-motion* digitalni video i kontrola putem prepoznavanja govora, kao i različite vrste podataka. Ovo je uslovilo postojanje novih servisa kao što su pristupi multimedijalnim bazama podataka (audio i video), kao i postojanje inteligentnih mreža koje omogućavaju komunikaciju sa ovim servisima ili sa drugim ljudima koji se nalaze na bilo kom mestu i u bilo koje vreme. Iz ovoga je očigledna potreba da se fiksne radne stanice zamene prenosivim koje će se odlikovati sa što manjom potrošnjom energije.

Čak i u neprenosivim aplikacijama, problem niske potrošnje postaje veoma značajan. Ranije ovaj problem nije bio tako značajan, obzirom da su korištena velika kućišta za čipove, a disipirana energija je odvođena različitim vrstama hladnjaka. Ipak, sa povećanjem gustine i veličine čipova i sistema, poteškoće u obezbeđivanju adekvatnog hlađenja mogu značajno povećati cenu sistema ili ograničiti njegovu funkcionalnost.

Stoga, evidentna je potreba za digitalnim sistemima koji se odlikuju malom potrošnjom i velikom propusnošću. Na sreću, prisutni su jasni tehnološki trendovi koji dopuštaju određeni stepen slobode, tako da je moguće udovoljiti svim ovim kontradiktornim zahtevima. Smanjivanje veličine komponenata, zajedno sa razvojem nisko-parazitnih pakovanja velike gustine, kao što su multičip moduli, umanjuju zabrinutost koja se odnosi na broj tranzistora koji se koriste. Za ilustraciju, 2 µm MOS tehnologija omogućava smeštanje od 1 do 10 x 10^9 tranzistora na prostoru od 8 in x 10 in ukoliko se koristi tehnologija veoma gustog pakovanja. Postavlja se pitanje kako će se ovo povećanje odraziti na potrošnju i funkcionisanje pri maloj potrošnji. Prethodna analiza koja se bavi pitanjem najboljeg iskorišćenja gustine tranzistora na nivou čipa, ukazuje na zaključak da je za mikroprocesore visokih performansi, najbolje obezbediti veliku količinu

memorije na čipu. Biće pokazano da je za funkcije koje zahtevaju intenzivno izračunavanje najbolje obezbediti dodatna kola za paralelno izvršavanje.

Drugo bitno razmatranje, posebno kod prenosivih aplikacija, odnosi se na izvršavanje u realnom vremenu (radio modem, kompresija video i govornog signala, prepoznavanje govora) i zahteva izračunavanje koje je uvek blisko maksimalnim brzinama. Standardne šeme za očuvanje energije u laptop računarima, koje su uglavnom zasnovane na *power-down* šemama, nisu podesne za ovakva, kontinualno aktivna, izračunavanja. S druge strane, postoji određeni stepen slobode kada je reč o implementaciji ovakvih funkcija. Naime, u trenutku kada se zahtevi za obradom u realnom vremenu jedanput ispune, dalja povećanja propusnosti ne dovode do nikakvog poboljšanja. Ova činjenica, zajedno sa mogućnošću korišćenja "neograničenog" broja tranzistora, omogućava strategiju razvoja dizajna arhitekture, kojom se mogu postići značajne uštede u energiji.

II. IZVORI DISIPACIJE SNAGE

Postoje tri osnovna izvora disipacije snage u CMOS kolima, koja su sadržana u sledećoj jednačini:

$$P_{total} = p_t\left(C_L \cdot V \cdot V_{dd} \cdot f_{clk}\right) + I_{sc} \cdot V_{dd} + I_{leakage} \cdot V_{dd} \qquad (1)$$

Prvi član predstavlja komutacionu komponentu snage, gde je C_L kapacitivnost opterećenja, f_{clk} je taktna frekvencija, a p_t je verovatnoća da se desila promena stanja pri kojoj je došlo do utroška energije (aktivacioni faktor). U većini slučajeva promena napona V je jednaka naponu napajanja V_{dd} ; ipak u nekim logičkim kolima, kao npr. kod implementacije *single-gate pass* tranzistora, promena napona u nekim internim tačkama može biti nešto manja. Drugi član potiče od direktne struje kratkog spoja I_{sc}, koja raste u slučaju kada su NMOS i PMOS tranzistori istovremeno aktivni pri čemu struja teče direktno od napajanja ka masi. Konačno, struja curenja, $I_{leakage}$, koja se javlja usled injekcije u supstrat i podpragovskih efekata, primarno je određena karakteristikama tehnologije. Dominantan član u jednom "*well-designed*" kolu preedstavlja komutaciona komponenta, pa se tako *low-power* dizajn svodi na minimizaciju p_t, C_L, V_{dd} i f_{clk}, pri čemu je potrebno očuvati zahtevanu funkcionalnost.

Proizvod snaga-kašnjenje može se interpretirati kao količina energije utrošena pri svakoj komutaciji (ili promeni) i kao takav posebno je koristan za poređenje disipacija snage kod elektronskih kola napravljenih različitim postupcima. Ako se pretpostavi da je bitna samo komutaciona komponenta disipacije snage, tada se može pisati

$$\text{energija po promeni} = P_{total} / f_{clk} = C_{effective} V_{dd}^2 \qquad (2)$$

gde je $C_{effective}$ efektivna kapacitivnost komutovana za izvršenje izračunavanja i data je sa $C_{effective} = p_t \cdot C_L$.

III. DIZAJN KOLA I TEHNOLOGIJA

Posotoji nekoliko opcija pri izboru osnovnog pristupa kolu i topologiji za implementiranje različitih logičkih i aritmetičkih funkcija. Izbor između statičkih i dinamičkih implementacija, *low-pass* i konvencionalnog CMOS stila, kao i između sinhronog i asinhronog tajminga samo su neke od opcija koje se nude projektantu sistema. Na drugom nivou, takođe postoje različite arhitekturno/strukturne opcije za implementiranje date logičke funkcije; npr. za implementiranje jednog sabirača može se koristiti *ripple-carry*, *carry-select* ili *carry-lookahead* topologija. U

ovom poglavlju biće razmatrane različite opcije za *low-power* koje nude određeni pristupi projektovanju elektronskih kola, pri čemu će pažnja biti usmerena i na neke opšte stvari i faktore koji utiču na izbor familije logičkih kola.

A. Dinamička u odnosu na statička logička kola

Izbor između statičkih i dinamičkih logičkih kola, pored njihovih *low-power* performansi, zavisi i od mnogih drugih faktora, kao što su npr. testabilnost i jednostavnost izrade. Ipak, kada je reč samo o *low-power* performansama, uočava se da dinamička logička kola imaju neke bitne prednosti kada je reč o broju prostornih jedinica, uključujući redukovanu komutacionu aktivnost usled hazarda i eleminaciju disipacije na kratkim spojevima. Statička logička kola imaju prednost jer ne postoji *precharge* operisanje i ne postoji raspodela naelektrisanja. Ova razmatranja biće detaljnije razmotrena:

1) *Lažne promene stanja*: Kod statičkih kola mogu se manifestovati lažne promene stanja usled konačnih propagacionih kašnjenja od jednog logičkog bloka do sledećeg (takođe se zovu i kritični putevi i dinamički hazardi), tj. u jednoj tački mogu se desiti različite promene za vreme jednog taktnog intervala pre nego što se ona postavi na pravu vrednost. Na primer, razmatrajmo jedan statički N-bitni sabirač, kod koga svi bitovi sabiraka prelaze sa logičke *nule* na *jedinicu*, sa bitom ulaznog prenosa postavljenim na *nulu*. Za sve bitove, rezultujuća suma treba da bude *nula*; ipak, propagacija signala prenosa uzrokuje da se na većini izlaza za kratko pojavi *jedinica*. Ova lažna promena disipira dodatnu energiju u odnosu na onu koja je potrebna za izvršenje izračunavanja. Broj ekstra promena predstavlja funkciju broja ulaznih kombinacija, internih dodela stanja unutar logike, kašnjenja košenja signala, i logičke dubine. Da bi se ilustrovala ova pojava, recimo da 8-bitni *ripple-carry* sabirač sa uniformno rasopodeljenim skupom slučajnih ulaznih kombinacija, utroši 30% ekstra energije. Iako je pažljivim projektovanjem moguće eliminisati ovaj problem, kod dinamičke logike ovakav problem ne postoji, obzirom da svaka tačka može imati samo jednu promenu u toku jednog taktnog intervala.

2) *Struje kratkih spojeva*: Struje kratkih spojeva (na direknoj stazi), I_{sc} u (1), prisutne su u statičkim CMOS kolima. Ipak, podešavanjem veličine tranzistora tako da vremena uspinjanja i opadanja signala budu jednaka, komponenta kratkog spoja ukupne disipacije snage može se održati ispod 20% (tipično < 5-10%) dinamičke komutacione komponente. Dinamička logika ne pokazuje ovu vrstu problema osim u onim slučajevima u kojima se koriste statičke *pull-up* komponente za kontrolu raspodele naelektrisanja ili kada je košenje signala značajno.

3) *Parazitna kapacitivnost*: Dinamička logička kola obično koriste manji broj tranzistora potrebnih za implementaciju date logičke funkcije, čime se direktno smanjuje količina kapacitivnosti koja komutira što utiče na proizvod snaga-kašnjenje (*power-delay)*. Ipak, dodatni tranzistori mogu biti potrebni kako bi osigurali da podela naelektrisanja ne rezultuje u pogrešnoj proceni.

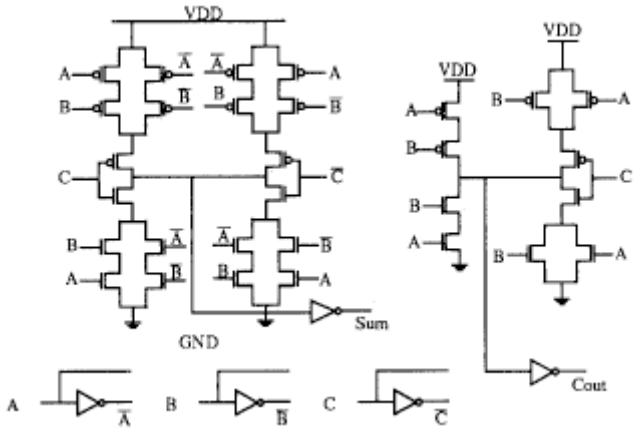
4) *Komutaciona aktivnost*: Dinamička logika ima značajan nedostatak kada je reč o operaciji prednaelektrisanja. Obzirom da kod dinamičke logike svaka tačka mora da bude prednaelektrisana u svakom taktnom intervalu, to znači da se neke tačke prednaelektrišu samo da bi ponovo bile momentalno ispražnjene prilikom određivanja vrednosti u njima, što vodi ka većem faktoru aktivnosti. Ukoliko jedan dvoulazni NOR gejt N tipa (prednaelektrisan na visok nivo) ima uniformnu ulaznu raspodelu visokih i niskih nivoa, tada će četiri moguće ulazne kombinacije (00, 01, 10, 11) biti podjednako verovatne. U tom slučaju postoji 75% verovatnoće da će ulazna tačka biti ispražnjena odmah nakon faze prednaelektrisanja, što implicira da je aktivnost za ovakav gejt 0.75 (tj. $P_{NOR} = C_L V_{dd}^2 f_{clk}$). S druge strane, faktor aktivnosti za statičku NOR komponentu biće samo 3/16, izuzimajući komponentu koja potiče od lažnih promena stanja. Snaga se vuče jedino prilikom NULA-u-JEDAN promene, pa je $p_{0\to1} = p(0)p(1) = p(0)(1-p(0))$. U opštem slučaju, aktivnosti gejtova će biti različite za statičku i dinamičku logiku i zavisiće od tipa operacije koja

se izvršava i verovatnoća ulaznih signala. U stvari, taktni baferi koji pobuđuju tranzistore za prednaelektrisanje takođe će zahtevati energiju koja kod statičke implementacije nije potrebna.

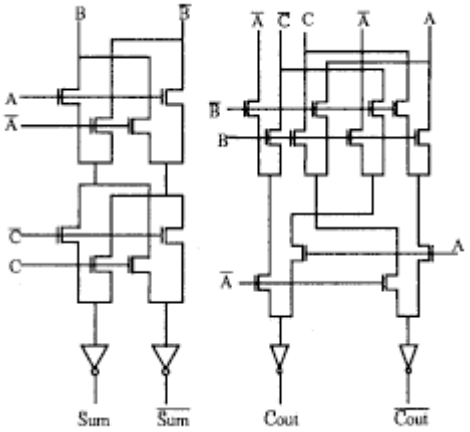
5) *Power-Down modovi*: U poslednje vreme, kod statičkih kola se efikasno koriste *power-down* tehnike zasnovane na zabrani taktnog signala, što nije pogodno za primenu kod dinamičkih tehnika. Ukoliko se žele sačuvati logička stanja za vreme isključenja, dinamičkim kolima je potrebno dodati relativno malo dodatnih kola, što rezultuje u blagom povećanju parazitne kapacitivnosti i smanjenoj brzini.

B. Konvencionalna statička logika u odnosu na pass-gate logiku

Mnogo je jasnija situacija kada je reč o korišćenju transfer gejtova za implementaciju logičkih funkcija, kakvi se koriste u familiji logičkih kola sa komplementarnim propusnim gejtovima (*CPL – Complementary Pass-gate Logic*). Na slici 1. prikazane su šeme tipičnog statičkog CMOS logičkog kola za puni brojač, zajedno sa odgovorajućom statičkom CPL verzijom. CPL dizajn koristi samo jednotransmisione NMOS gejtove umesto potpuno komplementarnih propusnih gejtova, kako bi redukovala kapacitivnost čvorova. *Pass-gate* logika je privlačna kada se zahteva što manji broj tranzistora za implementaciju bitnih logičkih funkcija, kao što je XOR koja zahteva samo dva propusna tranzistora u CPL implementaciji. Ova, naročito efikasna implementacija XOR je važna obzirom da predstavlja ključ za većinu aritmetičkih funkcija, omogućavajući da se sabirači i množači realizuju sa minimalnim brojem komponenata. Takođe, multiplekseri, registri i drugi ključni gradivni blokovi su pojednostavljeni korišćenjem *pass-gate* dizajna.



Broj tranzistora (konvencionalni CMOS) : 40



Broj tranzistora (CPL) : 28

Slika 1. Uporedjenje konvencionalnog CMOS i CPL sabirača

Ipak, CPL implementacija prikazana na slici 1 ima dva problema. Prvo, napon praga opada kroz jedno-kanalne propusne tranzistore, što rezultuje smanjenjem strujne pobude, a samim tim i sporijim radom pri sniženim naponima napajanja; ovo je važno za *low-power* dizajn obzirom da je poželjan rad na najnižim mogućim naponskim nivoima. Drugo, obzirom da visoki ulazni naponski nivo na regenerativnim invertorima nije V_{dd}, PMOS komponenta u invertoru nije potpuno isključena, tako da direktna statička disipavija snage može biti značajna. Da bi se rešili ovi problemi, pokazano je da je efikasno smanjenje napona praga, mada preveliko smanjenje dovodi do povećane disipacije usled podpragovskog curenja i redukovanih margina šuma. Pokazuje se da je disipacija snage kod sabirača iz *pass-gate* familije za 30% manja nego kod konvencionalnog statičkog dizajna, sa još značajnijom razlikom pri manjim naponima napajanja.

C. Skaliranje napona praga

Imajući u vidu da se sa korišćenjem niskopragovskih MOS komponenti mogu postići značajna poboljšanja u potrošnji, postavlja se pitanje do koje mere je moguće smanjivati napon praga. Granica je određena zahtevima za odgovarajuće margine šuma i povećanjem u podpragovskim strujama. Margine šuma će biti ublažene u *low-power* dizajnu zbog redukovanih struja, ali ipak, podpragovske struje mogu rezultovati u značajnoj statičkoj disipaciji snage. U opštem slučaju, podpragovske struje se javljaju usled difuzije nosilaca između sorsa i drejna kada napon između geta i sorsa V_{gs} dostigne tačku slabe inverzije, ali je još uvek ispod napona praga V_t, gde je dominantan drift nosilaca. U ovom režimu, MOSFET se ponaša približno bipolarnom tranzistoru, i podpragovska struja je eksponencijalno zavisna od napona između gejta i sorsa V_{gs}, i aproksimativno nezavisna od napona između drejna i sorsa V_{ds}, za V_{ds} približno veće od 0.1V. Sa ovim je povezan i podpragovski nagib $S_{th,}$ koji predstavlja vrednost napona potrebnu da podpragovska struja opadne za jednu dekadu. Na sobnoj temperaturi, tipična vrednost za S_{th} leži u granicama između 60 i 90 mV po dekadi, gde je 60 mV/dec donja granica. Jasno je da je manji nagib S_{th} bolji, obzirom da je poželjno da se komponenta isključi pri vrednost napona što bližoj V_t. Kao poređenje, za $L=1.5\ \mu m$, $W=70\ \mu m$ kod NMOS komponente, u tački u kojoj je V_{gs} jednako V_t, gde je V_t definisano za gustinu naelektrisanja površinske inverzije jednaku dopiranju supstrata, pokazano je približno curenje struje od 1μA, ili 0.014 μA/μm širine gejta. Od interesa je vrednost pri kojoj je ekstra struja nepoželjna u odnosu na prosečnu struju za vreme komutacije. Za CMOS invertor (PMOS: W = 8 μm, NMOS: W = 4 μm), izmerna je struja od 64μA u toku 3.7ns pri naponu napajanja od 2V. Ovo implicira 100% utrošak energije za podpragovsko curenje ukoliko bi komponenta radila na brzini takta od 25 MHz sa faktorom aktivnosti p_t = 1/6, tj. komponente su ostavljene u neaktivnom stanju i sa strujom curenja 83% vremena. Zbog toga nije preporučljivo koristiti komponente sa nultim pragom, već prag mora biti najmanje 0.2V, što obezbeđuje najmanje dva reda veličine smanjenje podpragovske struje. Ovo obezbeđuje dobar kompromis između poboljšanja pobudne struje i rada pri niskom naponu napajanja, pri čemu se podpragovska disipacija snage održava na zanemarljivom nivou. Ova vrednost može biti i veća u dinamičkim kolima kako bi se sprečilo slučajno pražnjenje za vreme faze očitavanja vrednosti. Na sreću, tehnolozi komponenata vide rešenje problema podpragovskih struja u budućim skaliranim tehnologijama, jer smanjenje napona napajanja utiče na smanjenje struje tako što redukuje maksimalni dozvoljeni drejn-sors napon. Dizajn budućih kola za *low-power* rad, trebao bi eksplicitno da uzme u obzir i efekat podpragovskih struja.

D. Power-Down strategije

U sinhronim kolima, logika između registara vrši kontinualno izračunavanje u toku svakog taktnog impulsa u zavisnosti od novih vrednosti na ulazima. Da bi se smanila potrošnja u

sinhronim kolima, važno je minimizovati komutacione aktivnosti isključivanjem (*powering down*) izvršnih jedinica dok one ne obavljaju korisne operacije. Ovo je veoma bitno jer logički moduli mogu da komutiraju i troše energiju čak i kada se oni aktivno ne koriste.

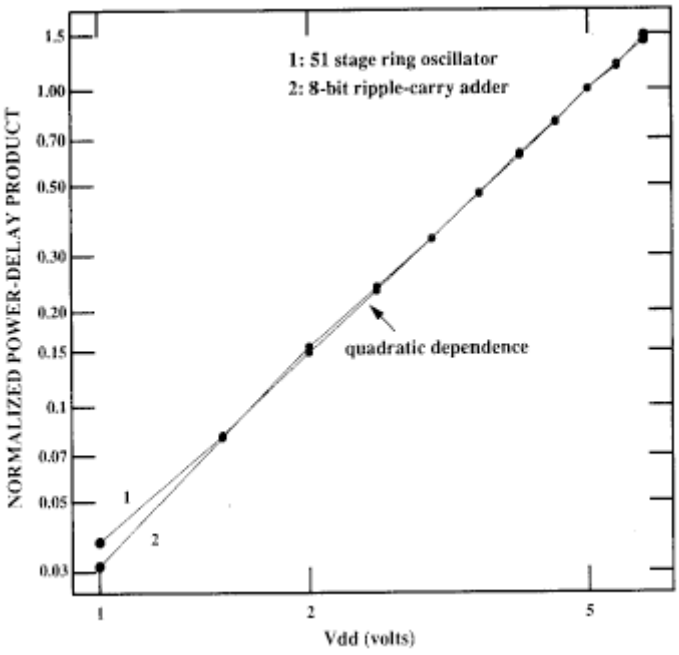
Dok dizajn sinhronih kola zahteva specijalne dizajnerske napore i *power-down* kola za detektovanje i isključivanje jedinica koje se ne koriste, samo-sinhronujuća logika ima inherentno isključivanje nekorišćenih modula, obzirom da se prelazi dešavaju samo kada su zahtevani. Ipak, obzirom da samo-sinhronujuća implementacija zahteva generisanje signala izvršenja koji pokazuje da su izlazi logičkog modula validni, potrebna se i dodatna kola. Postoji nekoliko pristupa za generisanje potrebnom signala izvršenja. Jedan metod je korišćenje *dual-rail* kodiranja, koje je kod određenih logičkih familija implicitno kao npr. DCVSL. Signal izvršenja u kombinacionoj makroćeliji napravljenoj kaskadnim vezivanjem DCVSL gejtova sastavljenih jednostavnim povezivanjem izlaza samo poslednjeg gejta preko ILI kola u lanac, vodi ka malim dodatnim zahtevima. Ipak, za svako izračunavanje, *dual-rail* kodiranje garantuje da će se komutacija izvršiti obzirom da najmanje jedan od izlaza mora da očita nulu. Pokzuje se da *dual-rail* DCVSL familija troši najmanje dva puta više energije po prelazu nego konvencionalna statička familija. Stoga, samo-sinhronujuće implementacije su se pokazale skupim u odnosu na energiju po stazama podataka koje neprestano vrše izračunavanja.

IV. SKALIRANJE NAPONA

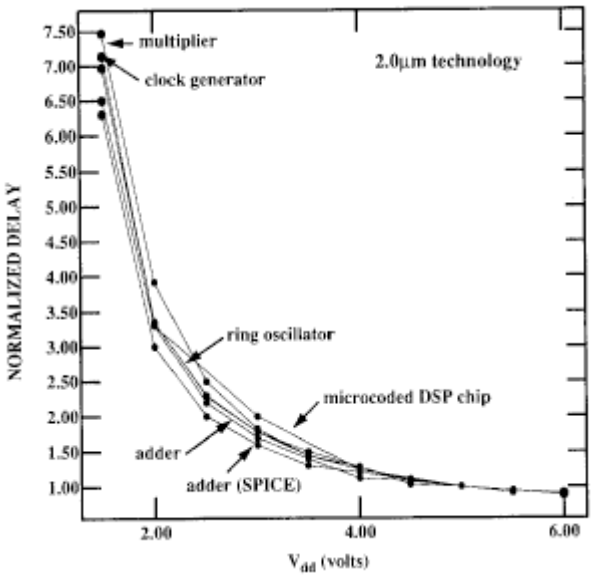
U dosadašnjem tekstu pažnja je prevashodno bila usmerena na udeo kapacitivnosti u izrazu CV^2f. Jasno je, takođe, da će smanjenje V doprineti još većim uštedama; zaista, redukovanje napona napajanja predstavlja ključnu tsvar za smanjenje potrošnje, čak i kada se uzmu u obzir modifikacije arhitekture sistema, koje su potrebne da bi se postigla odgovarajuća propusnost. Prvo, biće prikazano ponašanje kola (karakteristike kašnjenja i energije) u funkciji napona napajanja i veličina oblika strukture. U poređenju sa eksperimentalnim podacima, pokazano je da jednostavna teorija prvog reda daje zadivljujuće tačne reprezentacije različitih zavisnosti za širok spektar različitih kola i arhitektura. Zatim će biti razmotrena dva prethodna pristupa za skaliranje napona napajanja, gde je pažnja usmerena na postizanje odgovarajuće pouzdanosti i performansi. Ovo je praćeno arhitekturnim pristupom, iz koga su izvedeni “optimalni” napon napajanja zasnovan na tehnologiji, arhitekturi i ograničenjima koje postavljaju margine šuma.

A. Kašnjenje i proizvod snaga-kašnjenje (power-delay)

Kao što je naznačeno u (2), energija po promeni ili ekvivalentni proizvod snaga-kašnjenje u “odgovarajuće dizajniranim” CMOS kolima (o čemu je bilo reči u poglavlju II) je proporcionalna V^2. Ovo se vidi sa slike 2. koja prikazuje dva eksperimentalna koja pokazuju očekivanu V^2 zavisnost. Stoga, neophodno je samo redukovati napon napajanja za *kvadratno* poboljšanje u proizvodu snaga-kašnjenje za određenu logičku familiju.



Slika 2. Proizvod potrošnja-kašnjenje karakteriše kvadratna zavisnost za dva različita kola



Slika 3. Kašnjenje u funkciji napona napajanja za različita kola

Nažalost, ovo jednostavno rešenje za *low-power* dizajn snosi određene troškove. Na slici 3. je prikazan efekat redukovanja V_{dd} za različita logička kola različitih funkcija čija veličina varira od 56 do 44000 tranzistora; sva kola pokazuju suštinski istu zavisnost (tabela I).

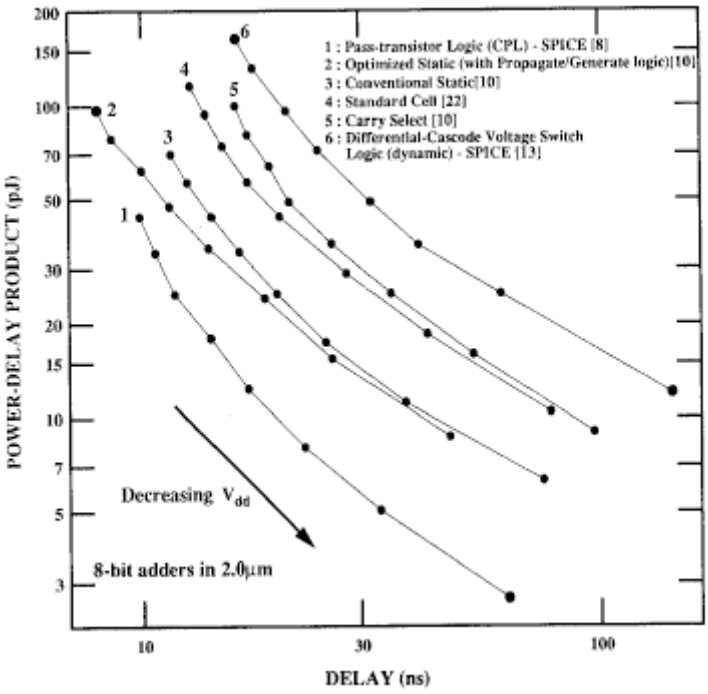
Tabela I
Detalji komponenata koji su prikazani na slici 3

Component (all in 2 μm)	# of Transistors	Area	Comments
Microcoded DSP Chip [21]	44 802	94 mm^2	20-b data path
Multiplier	20 432	12.2 mm^2	24 × 24 b
Adder	256	0.083 mm^2	conventional static
Ring Oscillator	102	0.055 mm^2	51 stages
Clock Generator	56	0.04 mm^2	cross-coupled NOR

Jasno je da se za smanjenje V_{dd} plaća smanjenjem brzine, pri čemu se kašnjenje drastično povećava kako se V_{dd} približava sumi napona pragova komponenata. Iako je tačna analiza kašnjenja prilično kompleksna ukoliko se uzmu u obzir nelinearne karakteristike CMOS gejtova, pokazuje se da derivacija prvog reda adekvatno predviđa eksperimentalno određenu zavisnost i data je sa

$$T_d = \frac{C_L \times V_{dd}}{I} = \frac{C_L \times V_{dd}}{\mu C_{ox}\left(w/L\right)\left(V_{dd} - V_t\right)^2}$$

Takođe su izračunate (eksperimentalnim merenjima i SPICE simulacijama) karakteristike za energiju i kašnjenje kod nekoliko različitih logičkih stilova i topologija korišćenjem 8-bitnog sabirača kao referentnog; rezultati su prikazani na log-log dijagramu sa slike 4.



Slika 4. Proizvod snaga disipacije-kašnjenje u funkciji brzine rada za različite tipove kola

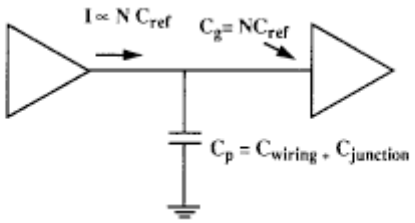
Uočava se da se proizvod snaga-kašnjenje povećava sa kašnjenjem (kroz redukciju napona napajanja), pa je stoga poželjno da se radi na najmanjoj mogućoj brzini. Obzirom da je cilj da se

redukuje potrošnja energije uz očuvanje zadovoljavajuće propusnosti, potrebna je kompenzacija za ova povećana kašnjenja na niskim naponima. Od posebnog je interesa opseg energija potrebnih za prelaz pri datom kašnjenju. Najbolja logička familija koja je bila analizirana (10 puta bolja od najgore ispitivane) bila je *pass-gate* familija, CPL.

Slike 2,3, i 4 sugerišu da je ponašanje kašnjenja i energije u funkciji V_{dd} skaliranja za datu tehnologiju "prihvatljivo" i relativno nezavisno od logičkog stila i kompleksnosti kola. Ovi rezultati su iskorišćeni za optimizaciju arhitekture za malu potrošnju, tretiranjem V_{dd} kao slobodne promenjive i dozvoljavanjem arhitekturama da variraju kako bi zadržale konstantnu propusnost. Korišćenjem monotonih zavisnosti kašnjenja i energije od napona napajanja za različite varijacije kola, moguće je ostvariti relativno jaka predviđanja o tipovima arhitektura koje su najbole za *low-power* dizajn. Naravno, kao što je već ranije spomenuto, postoje neki logički stilovi kao što je NMOS *pass-transistor logic* bez redukovanih napona pragova čije energetske i karakterisitike kašnjenja odstupaju od ovih slučajeva, pa su i kvantitavni rezultati drugačiji, ali osnovni zaključci i dalje važe.

B. Optimalna veličina tranzistora sa skaliranjem napona

Nezavisno od izbora logičke familije i topologije, određivanje optimalne veličine tranzistora igra bitnu ulogu u redukovanju potrošnje. Za malu potrošnju, kada se govori o brzim kolima, važno je izjednačiti sve puteve kašnjenja kako pojedinačni kritični put ne bi ograničavao performanse celog kola. Ipak, pored ovih ograničenja, postavlja se pitanje u kom stepenu je potrebno kod svih komponenata uniformno povećavati odnos *W/L*, što doprinosi uniformnom smanjivanju kašnjenja gejta, kako bi se postiglo odgovarajuće smanjenje napona i snage. Pokazano je da ukoliko se dopusti da napon varira, optimalna veličina za rad na maloj potrošnji je sasvim različita od one za velike brzine.



Slika 5. Model kola kod analize efekta promene veličine tranzistora

Na slici 5 prikazano je jedno jednostavno kolo sa dva gejta, sa prvim stepenom koji pobuđuje kapcitivnost drugog, zajedno sa parazitnom kapacitivnošću C_p usled sprege sa supstratom i veze. Pretpostavljajući da je ulazna kapacitivnost gejta za oba stepena data sa NC_{ref}, gde C_{ref} predstavlja kapacitivnost gejta MOS komponentesa najmanjim dozvoljenim *W/L*, kašnjenje kroz prvi gejt pri naponu napajanja V_{ref} dato je sa

$$T_N = K\frac{\left(C_p + NC_{ref}\right)}{\left(NC_{ref}\right)}\frac{V_{ref}}{\left(V_{ref} - V_t\right)^2} = K\left(1+\alpha/N\right)\frac{V_{ref}}{\left(V_{ref} - V_t\right)^2}$$

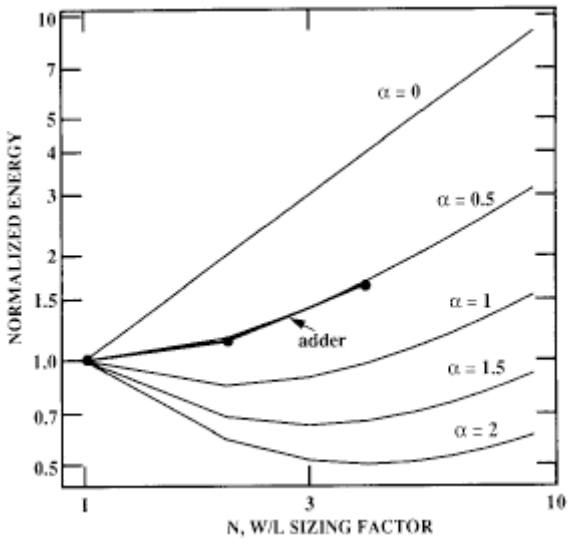
gde je α kao odnos C_p i C_{ref}, a K predstavlja članove nezavisne od širine komponente i napona. Za dati napon napajanja V_{ref}, ubrzanje kola čiji je odnos *W/L* smanjen za faktor N u odnosu na referentno kolo korišćenjem tranzistora minimalne veličine (N = 1) dato je sa (1+α/N)/(1+α). Da bi se procenile performanse dva dizajna iste brzine sa stanovišta energije, naponu skalirane solucije je dozvoljeno da varira kako bi se kašnjenje održalo konstantnim. Pretpostavljajući da se kašnjenje skalira kao $1/V_{dd}$ (ignorišući smanjenja napona praga u promenama napona), napon napajanja V_N, pri jednakim kašnjenjima skaliranog i referentnog dizajna, dato je sa

$$V_N = \frac{(1+\alpha/N)}{(1+\alpha)} V_{ref}$$

Pod ovim uslovima, energija koju troši prvi stepen u funkciji N data je sa:

$$Energy(N) = (C_p + NC_{ref})V_N^2 = \frac{NC_{ref}(1+\alpha/N)^3 V_{ref}^2}{(1+\alpha)^2}$$

Nakon normalizovanja sa E_{ref} (energija za slučaj minimalne veličine), slika 6 pokazuje *Energy(N)/Energy(1)* u odnosu na *N* za različite vrednosti α. Kada nema doprinosa parazitne kapacitivnosti (tj. $\alpha = 0$), energija se povećava linearno sa *N*, pa rešenja koja koriste komponente sa najmanjim *W/L* imaju najmanju potrošnju. Pri velikim vrednostima α, kada parazitne kapacitivnosti počnu da dominiraju u odnosu na kapacitivnosti gejta, energija se smanjuje privremeno sa povećanjem veličine komponenata a onda počinje da se povećava, rezultujući optimalnom vrednošću za *N*. Inicijalno smanjenje u naponu napajanja postignuto redukcijom kašnjenja više nego kompenzuje povećanje kapacitivnosti zbog povećanja *N*. Ipak, posle određene tačke povećanje kapacitivnosti dominira u odnosu na postignutu redukciju napona, obzirom da je povećanje inkrementalne brzine sa smanjenjem dimenzija tranzistora veoma malo (ovo se vidi iz (4), gde kašnjenje postaje nezavisno od α sa približavanjem N beskonačnosti). U analizi je pretpostavljeno da je parazitna kapacitivnost nezavisna od promena dimenzija komponenti. Ipak, difuzije u drejnu i sorsu i parametar kapacitivnosti se stvarno povećavaju sa povećanjem oblasti difuzija, čime se favorizuju manje dimenzije komponenata čineći prethodnu analizu, analizom najgoreg slučaja.



Slika 6. Energija u odnosu na faktor obima tranzistora za različite parazitne efekte

Takođe, na slici 6 su prikazani rezultati simulacije za ekstraktovane oblike 8-bitnog *carry* brojača za komponente sa tri različita odnosa *W/L* (N=1, N=2, i N=4). Kriva veoma dobro prati pojednostavljeni model prvog reda, sugerišući da je u ovom primeru dominantniji efekat kapacitivnosti gejta od efekta parazitne kapacitivnosti. U ovom slučaju, povećanjem *W/L* kod komponenata nije od pomoći, pa rešenje koje koristi najmanji mogući odnos *W/L* daje najbolje rezultate.

Iz ovog poglavlja jasno je da je određivanje “optimalnog” napona napajanja ključ za minimizaciju potrošnje, pa se u daljem tekstu razmatra ovaj efekat.

C. Skaliranje napona u odnosu na pouzdanost

Jedan od pristupa prilikom izbora optimalnog napona napajanja za duboke-submikrometarske tehnologije zasnovan je na optimiziranju kompromisa između brzine i pouzdanosti. Skaliranje konstantnog napona, najčešće korišćena tehnika, rezultuje u višim električnim poljima koja stvaraju vruće nosioce. Kao rezultat ovoga, komponenta se vremenom degradira (uključujući promene u naponima praga, degradaciju transkonduktanse, i povećanje subpragovskih struja), što vodi ka eventualnom otkazu. Jedno od rešenja za redukovanje vrućih nosilaca jeste promena fizičke strukture komponente, kao što je korišćenje slabo dopiranog drejna (LDD-*Lightly Doped Drain*), obično po cenu smanjenja performansi. Pretpostavljajući korišćenje 0.25μm tehnologiju izabiranjem tačke minimuma na krivoj kojaj pokazuje zavisnost kašnjenja od V_{dd}. Za napone iznad ove tačke minimuma, nađeno je da se kašnjenje povećava sa povećanjem V_{dd}, obzirom da LDD struktura korišćena iz razloga pouzdanosti rezultuje u povećanim parazitnim otpornostima.

D. Skaliranje napona u odnosu na tehnologiju

Pojednostavljena analiza kašnjenja prvog reda prikazana u odeljku IV-A je razumljivo tačna za dugo-kanalne komponente. Ipak, sa smanjenjem veličina ispod 1.0 μm, karakteristike kašnjenja u funkciji smanjenja napona napajanja odstupaju od prezentovane teorije prvog reda obzirom da ne uzimaju u obzir zasićenje brzine nosilaca pod visokim električnim poljem. Kao rezultat zasićenja brzine, struja više nije kvadratna funkcija napona već linearna; zbog toga, struja je značajno redukovana i aproksimativno je data sa $I = WC_{ox}\left(V_{dd} - V_t\right)v_{\max}$. Odavde i iz jednačine za kašnjenje u (3), vidimo da je kašnjenje submikronskih kola relativno nezavisno od napona napajanja pri visokim električnim poljima.

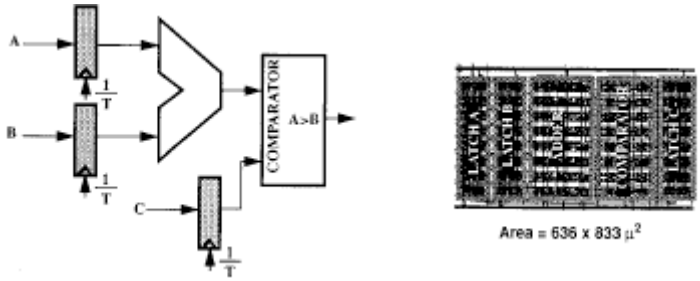
Pristup zasnovan na tehnologiji predlaže da se u submikronskim tehnologijama odabere napon napajanja kojim se održavaju performanse brzine. Korišćenjem relativne nezavisnosti kašnjenja od napona napajanja na visokim električnim poljima, napon se može spustiti do izvesne mere kod komponente sa zasićenjem brzine, sa veoma malim uticajem na performanse brzine. Ovo pokazuje da je dobitak iznad određene vrednosti napona veoma mali. Ova ideja je formalizovana od strane *Kakumu-a* i *Kinugawa*-e, doprinoseći konceptu "kritičnog napona" koji obezbeđuje donju granicu napona napajanja. Kritični napon je definisan kao $V_c = 1.1E_cL_{eff}$, gde je E_c kritično električno polje koje uzrokuje zasićenje brzine; ovo je napon pri kome se kriva zavisnosti kašnjenja od V_{dd} približava $\sqrt{V_{dd}}$ zavisnosti. Za 0.3μm tehnologiju, nalazi se da je predložena donja granica napona napajanja (ili kritični napon) 2.43V.

Zbog ovog efekta, prelaskom na 3.3V industrijski standard postignuto je 60% smanjenje potrošnje, pri čemu nisu značajno degradirane performanse brzine.

E. Skaliranje napona u odnosu na arhitekturu

Gore spomenuti "tehnološki" pristupi bili su fokusirani na redukovanje napona uz očuvanje brzine komponente, i nisu pokušavali da postignu minimalnu moguću energiju. Kao što je prikazano na slikama 2 i 4, CMOS logički gejtovi postižu niže vrednosti proizvoda kašnjenje-snaga (energija po komutaciji) sa smanjenjem napona napajanja. U stvari, kada se komponenta nađe u stanju zasićenja brzine dolazi do dalje degradacije u energiji po izračunavanju, pa minimiziranjem energije potrebne za izračunavanje, *Kakumu*-ov i *Kinagawa*-in kritični napon obezbeđuje gornju granicu napona napajanja (dok je za njihovu analizu obezbedio nižu granicu).

Sada je zadatak arhitekture da kompenzuje smanjenu brzinu kola koja je posledica rada ispod kritičnog napona.

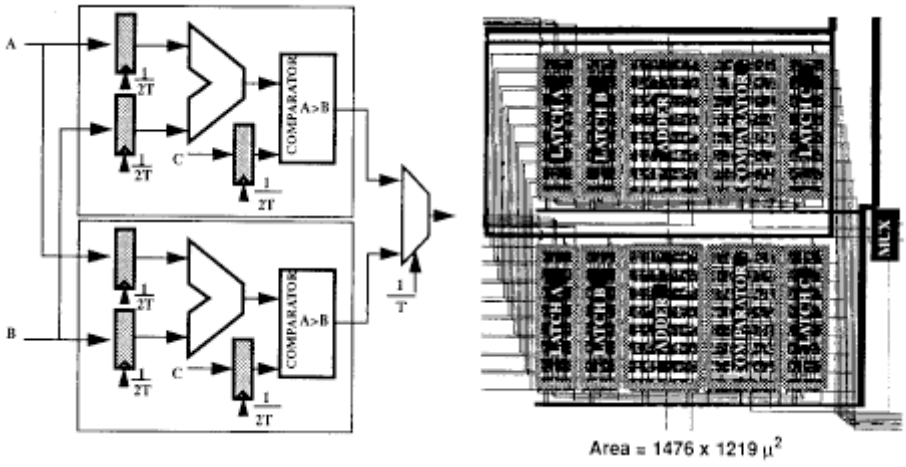


Slika 7. Jednostavna staza podataka i odgovarajući *layout*

Da bi se ilustrovalo kako arhitekturne tehnike mogu biti iskorišćene za kompenzaciju redukovanih brzina, analizirana je jedna jednostavna 8-bitna staza podataka koja sadrži sabirač i komarator, pri čemu je pretpostavka da se radi o 2.0μm tehnologiji. Kao što je prikazano na slici 7, ulazi *A* i *B* se sabiraju, a rezultat se upoređuje sa ulazom *C*. Pretpostavljajući najgori slučaj, kašnjenje kroz sabirač, komparator i leč je približno 25ns pri naponu napajanja 5V, pa sistem u najboljem slučaju može biti taktovan signalom periode T = 25ns. Kada se zahteva rad sa ovom maksimalnom proposnošću, jasno je da radni napon ne može biti više redukovan jer dodatno kašnjenje ne može biti tolerisano, tako da ne dolazi do smanjenja potrošnje. Ova staza podataka biće korišćena kao referentna u daljoj analizi sa stanovišta arhitekture i sva prezentovane vrednosti za poboljšanje smanjenja potrošnje odnosiće se na nju. Snaga referentne staze podataka data je sa:

$$P_{ref} = C_{ref} V_{ref}^2 f_{ref}$$

gde je C_{ref} ukupna efektivna kapacitivnost komutirana za jedan taktni interval. Efektivna kapacitivnost je određena usrednjavanjem energije za sekvencu ulaznih kombinacija sa uniformnom raspodelom.



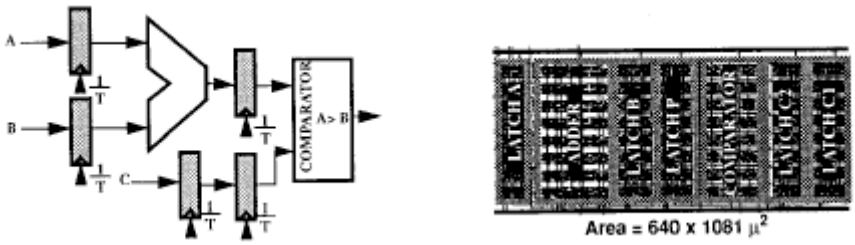
Slika 8. Paralelena implementacija jednostavne staze podataka

Jedan od načina da se održi propusnost prilikom smanjenja napona napajanja jeste korišćenje paralelne arhitekture. Kao što je prikazano na slici 8, koriste se dve identične sabirač-komparator staze podataka, pri čemu je svakoj jedinici dozvoljeno da radi na polovini originalne brzine pri čemu se održava originalna propusnost. Obzirom da su se brzinski zahtevi za sabirač, komparator i leč smanjili sa 25 na 50ns, napon može biti smanjen sa 5 na 2.9V (napon pri kome je kašnjenje udvostručeno, sa slike 3). Dok je kapacitivnost staze podataka povećana faktorom 2,

radna frekvencija je shodno tome smanjena faktorom 2. Na žalost, postoji i blago povećanje ukupne "efektivne" kapacitivnosti koje se javlja zbog dodatnog rutiranja, što rezultuje u povećanju kapacitivnosti faktorom 2.15. Tako je snaga za paralelnu stazu podataka data sa

$$P_{par} = C_{par} V_{par}^2 f_{par} = \left(2.15 C_{ref}\right)\left(0.58 V_{ref}\right)^2 \left(\frac{f_{ref}}{2}\right) \approx 0.36 P_{ref}$$

Ovaj metod redukovanja snage korišćenjem paralelizma ima za posledicu povećanje površine, i nije pogodan za prostorom-ograničene dizajne. U opštem slučaju, paralelizam će imati za posledicu dodatno rutiranje (a samim tim i dodatnu potrošnju), pa se mora izvršiti pažljiva optimizacija kako bi se minimizovao ovaj nedostatak (npr. tehnike particionisanja za minimalne gubitke). Kapacitivnost veza će takođe igrati bitnu ulogu u duboko-submikrometarskim implementacijama, obzirom da komponenta ivične kapacitivnost u kapcitivnosti veza (C_{wiring} = C_{area} + $C_{fringing}$ + C_{wiring}) može postati dominantan član u ukupnoj kapacitivnosti (jednakoj C_{gate} + $C_{junction}$ + C_{wiring}) i sprečiti skaliranje.



Slika 9. Protočna implementacija jednostavne staze podataka

Drugi mogući pristup predstavlja primena protočne strukture, kao što je prikazano na slici 9. Sa dodatnim protočnim lečom, kritični put postaje max[T_{adder},$T_{comparator}$], što omogućava sabiraču i komparatoru da rade na manjoj brzini. U ovom primeru, dva kašnjenja su jednaka, što ponovo omogućava naponu napajanja da se smanji sa referentnih 5V na 2.9V (napon pri kome se kašnjenje udvostručuje) bez smanjenja propusnosti. Ipak, ovom tehnikom je smanjen gubitak prostora obzirom da je potrebno dodati samo protočne registre. Primećuje se da ponovo postoji slabo povećanje hardvera usled dodatnih lečeva, što povećava "efektivnu" kapacitivnost približno faktorom 1.15. Potrošnja protočne putanje podataka je

$$P_{pipe} = C_{pipe} V_{pipe}^2 f_{pipe} = \left(1.15 C_{ref}\right)\left(0.58 V_{ref}\right)^2 f_{ref} \approx 0.39 P_{ref}$$

Sa ovom arhitekturom, potrošnja se smanjuje za faktor približno 2.5, dajući približno isto smanjenje potrošnje kao i u slučaju paralelizma sa prednošću zbog manjeg zauzeća prostora. Pored toga, povećanje nivoa protočnosti takođe ima efekat u smanjenju logičke dubine a samim tim i potrošnje usled hazarda i kritičnih puteva.

Takođe, očigledno je i poboljšanje koje se dobija kombinacijom protočnosti i paralelizma. Obzirom da ova arhitektura redukuje kritični put, a samim tim i brzinske zahteve za faktor 4, napon može biti smanjen sve do povećanja kašnjenja za faktor 4. Potrošnja je u ovom slučaju

$$P_{parpipe} = C_{parpipe} V_{parpipe}^2 f_{parpipe} = \left(2.5 C_{ref}\right)\left(0.4 V_{ref}\right)^2 \left(\frac{f_{ref}}{2}\right) \approx 0.2 P_{ref}$$

Paralelno-protočna implementacija rezultuje u petostrukom smanjenju potrošnje. Tabela II pokazuje uporedne parametre za različite arhitekture opisane za jednostavnu sabirač-komparator putanju podataka.

Tabela II
Sumarni pregled osobina različitih arhitektura

Architecture Type	Voltage	Area	Power
Simple data path (no pipelining or parallelism)	5 V	1	1
Pipelined data path	2.9 V	1.3	0.39
Parallel data path	2.9 V	3.4	0.36
Pipeline parallel	2.0 V	3.7	0.2

Iz prethodnih primera, jasno je da su klasične vremenski-multipleksirane arhitekture, kakve se koriste u mikroprocesorima opšte namene i DSP čipovima, najmanje poželjne za *low-power* primene. Ovo sledi iz činjenice da multipleksiranje povećava brzinske zahteve u kolima, ne dozvoljavajući tako smanjenje u naponu napajanja.

V. OPTIMALNI NAPON NAPAJANJA

U prethodnim odeljcima, pokazano je da povećanje kašnjenja usled smanjenja napona napajanja može biti kompenzovano korišćenjem paralelnih arhitektura. Ipak, kao što se vidi sa slike 3. i (3), kako se napon napajanja približava naponima praga komponenata, kašnjenja gejta se ubrzano povećavaju. Analogno tome, stepen paralelizma i dodatnih kola raste do tačke u kojoj povećanje obima kola dominira nad smanjenjem potrošnje usled daljeg smanjenja napona, što dovodi do postojanja "optimalnog" napona sa stanovišta arhitekture. Za određivanje vrednosti ovog napona, koristi se sledeći model za potrošnju pri fiksiranoj propusnosti sistema u funkciji napona (odn. stepena paralelizma):

$$Power(N) = NC_{ref}V^2\frac{f_{ref}}{N} + C_{ip}V^2\frac{f_{ref}}{N} + C_{interface}V^2 f_{ref}$$

gde je N broj paralelnih procesora, C_{ref} kapacitivnost pojedinačnog procesora, C_{ip} kapcitivnost usled međuprocesorske komunikacije koja se javlja zbog paralelizma (usled kontrole i rutiranja), i $C_{interface}$ je dodatna kapacitivnost koja se javlja u interfejsu u kome se sa uvođenjem paralelnije arhitekture ne smanjuje brzina. U opštem slučaju, C_{ip} i $C_{interface}$ su funkcije od N, pa poboljšanje potrošnje u odnosu na referentni slučaj (tj. bez paralelizma) može da se iskaže kao:

$$P_{normalized} = \left(1 + \frac{C_{ip}(N)}{NC_{ref}} + \frac{C_{interface}(N)}{C_{ref}}\right)\left(\frac{V}{V_{ref}}\right)^2$$

Pri veoma niskim naponima napajanja (bliskim naponima praga komponenata), broj procesora (i odgovarajuće povećanje površine) raste brže nego što se član V^2 smanjuje, što rezultuje u povećanju potrošnje sa daljim smanjenjem napona.

Komponente sa smanjenim naponima praga imaju tendenciju da smanje optimalni napon; ipak, kao što se vidi u odeljku III-C, pri naponima praga ispod 0.2V, disipacija snage usled subpragovske struje počinje da dominira i ograničava dalje poboljšanje potrošnje. Nađeno je da je donja granica napona napajanja kod CMOS invertora sa "ispravnom" funkcionalnošću 0.2V ($8kT/q$). Ovo daje granicu za proizvod snaga-kašnjenje koji se može postići u CMOS digitalnim

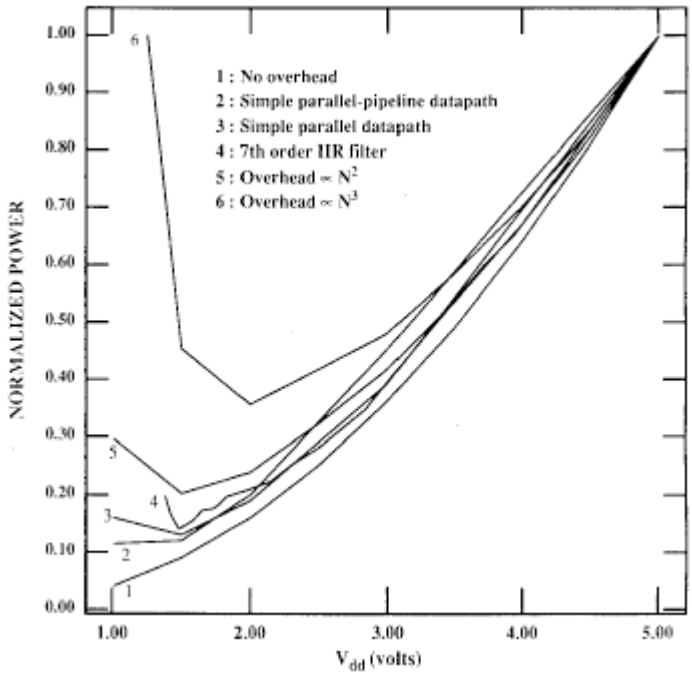
kolima; ipak, stepen paralelizma koji zadržava propusnost pri ovom nivou napona će bez sumnje postavljati ograničenja u svakoj konkretnoj situaciji.

U dosadašnjem tekstu, pokazano je da paralelne i protočne strukture mogu obezbediti smanjenje napona napajanja do optimalnog nivoa; ovo će i biti slučaj ukoliko implementirani algoritam ne prikazuje ponavljanje (povratnu spregu). Međutim, veliki je broj aplikacija koje su po prirodi prilično rekurzivne, počev od onih jednostavnih kao što su beskonačni impulsni odgovor i adaptivni filtri, do onih kompleksnijih kakvo je npr. sistemsko rešavanje nelinearnih jednačina i algoritmi za adaptivnu kompresiju. Zbog toga postoji i algoritamsko ograničenje u odnosu na nivo do koga paralelizam i protočnost mogu biti korišćeni za smanjenje napona. I pored toga što korišćenje transformacija u grafovima kontrolnih podataka može da do izvesne mere prevaziđe ovo usko grlo, smanjenje napona do optimalne vrednosti uslovljeno je i ograničenjima vezanim za kašnjenje i strukturom izračunavanja pojedinih algoritama.

Još jedno ograničenje za najniži dozvoljeni napon napajanja predstavlja sistemska margina šuma ($V_{noise\ margin}$). Zbog toga je potrebno ograničiti optimalni napon sa donje strane, pa je

$$V_{noise margin} \leq V_{optimal} \leq V_{critical}$$

sa $V_{critical}$ definisanim u poglavlju IV-D. Zbog ovoga će "optimalni" napon napajanja (za određenu tehnologiju) ležati negde između napona koga diktira margina šuma i kritičnog napona.



Slika 10. Optimalni radni napon

Slika 10. prikazuje optimalnu potrošnju (normalizovanu do 1 na V_{dd} = 5V) u funkciji V_{dd} za različite slučajeve u 2.0μm tehnologiji. Kao što će biti pokazano, postoji široka raznovrsnost pretpostavki u ovim različitim slučajevima i veoma je bitno da se primeti da sve one imaju, grubo posmatrano, istu optimalnu vrednost napona napajanja, približno 1.5V. Kriva 1 na ovoj slici prikazuje disipaciju snage koja bi se postigla ukoliko ne bi bilo povećanja površine usled povećanja stepena paralelizma. U ovom slučaju, snaga je strogo opadajuća funkcija od V_{dd} pa je optimalni napon postavljen na minimalnu vrednost koju određuje margina šuma (pretpostavljajući da se ne dostiže rekurzivno usko grlo). Kriva 5 pretpostavlja da unutarprocesorska kapacitivnost ima N^2 zavisnost dok kriva 6 pretpostavlja zavisnost N^3. Očekuje se da je u većini praktičnih slučajeva zavisnost manja od N^2, ali čak i sa ekstremno jakom N^3 zavisnošću nalazi se da je optimalna vrednost oko 2V.

Krive 2 i 3 su dobijene iz podataka za stvarne topologije, i prikazuju zavisnost od kapacitivnosti interfejsa koja se kreće između linearne i kvadratne zavisno od stepena paralelizma, *N*. Za ove slučajeve, nije bilo interprocesorske komunikacije. Krive 2 i 3 predstavljaju proširenja primera iz odeljka IV-E u kome su paralelne i paralelno-protočne implementacije jednostavne staze podataka duplicirane *N* puta. Kriva 4 prikazuje jedan mnogo komplikovaniji primer, IIR filtar sedmog reda, takođe dobijen iz podataka za stvarnu topologiju. Gubici u ovom slučaju potiču primarno od interprocesorske komunikacije. Ova kriva se prekida na oko 1.4V jer je u toj tački algoritam napravio maksimalni paralelizam, dostižući rekurzivno usko grlo. Za ovaj slučaj, pri naponu napajanja od 5V, arhitektura je u osnovi jedna hardverska jedinica koja je vremenski multipleksirana i koja zahteva oko 7 puta više energije nego optimalni paralelni slučaj koji je postignut sa napajanjem od oko 1.5V.

Tabela III
Normalizovana površina/disipacija za različiti napon karakteristike 2, 3, i 4 sa slike 10.

Voltage	Parallel Area/Power	Parallel-Pipeline Area/Power	IIR Area/Power
5	1/1	1/1	1/1
2	6/0.19	3.7/0.2	2.6/0.23
1.5	11/0.13	7/0.12	7/0.14
1.4	15/0.14	10/0.11	Recursive bottleneck reached

Tabela 3 sadrži smanjenja potrošnje i normalizovane površine koje su dobijene iz topologija. Povećanje u površini ukazuje na stepen paralelizma koji se koristi. Ključna stvar je u tome što se pokazuje da je optimalni napon relativno nezavistan u svim razmatranim slučajevima, i kreće se oko 1.5V za 2.0μm tehnologiju; slična analiza koja je koristila napon praga od 0.5V, za 0.8μm proces (sa L_{eff} 0.5μm) rezultovala je u optimalnom naponu od oko 1V, sa smanjenjem potrošnje za faktor 10. Dalje smanjivanje napona praga bi omogućilo rad i na nižim naponima, a samim tim i veće uštede energije.

VI. ZAKLJUČCI

Kada se radi o *low power* dizajnu, u obzir se uzima mnoštvo razmatranja koja uključuju logički stil, tehnologiju koja se koristi, i implementiranu logiku. Prikazani faktori koji doprinose disipaciji snage uključuju lažne promene stanja usled hazarda i uslova na kritičnom putu, struje curenja i direktnog puta, promene stanja izazvane prednaelektrisanjem i promene stanja u nekorišćenim kolima. Pokazano je da je logička familija propusnih gejtova (*pass-gate)* sa modifikovanim naponimam pragova najbolji izbor za *low-power* dizajn, pre svega zbog minimalnim brojem gejtova potrebnih za realizaciju bitnih logičkih funkcija. Analiza smanjenja dimenzija tranzistora je pokazala da tranzistore sa smanjenim dimenzijama treba koristiti kada su parazitne kapacitivnosti manje od kapacitivnosti aktivnih gejtova u kaskadi logičkih gejtova.

Sa postojećim trendom razvoja tehnologija kroz smanjenje dimenzija i poboljšanje tehnika pakovanja, omogućen je jedan novi stepen slobode u arhitekturnom dizajnu u kome površina silicijuma može biti kompenzovana za potrošnju. Paralelne arhitekture koje koriste protočnost ili repliciranje hardvera, obezbeđuju mehanizam za ovu kompenzaciju, održavajući propusnost koristeći sporije komponente omogućavajući tako rad pri smanjenom naponu. Priroda zavisnosti disipacije snage i kašnjenja u funkciji napona napajanja za različite situacije omogućava optimizacije arhitekture. Na ovaj način, pronađeno je da je za različite situacije optimalni napon manji od 1.5V, ispod koga povećanje površine povezano sa paralelizmom postaje dominantno.

Postoje i druga ograničenja koja ne dozvoljavaju postizanje optimalnog napona napajanja. Implementirani algoritam može biti sekvencijane prirode i/ili imati povratnu spregu koja će ograničiti stepen paralelizma koji može biti primenjen. Druga mogućnost je da optimalni stepen

paralelizma može biti tako veliki da broj tranzistora bude nerazumno veliki čineći optimalno rešenje nerezonskim. Ipak, u svakom slučaju, cilj prilikom minimizacije potrošnje je jasan: rad kola što je moguće sporije, sa najnižim mogućim naponom napajanja.