



УНИВЕРЗИТЕТ У НИШУ
ЕЛЕКТРОНСКИ ФАКУЛТЕТ



Бојан Јовановић

АНАЛИТИЧКИ МОДЕЛ ЗА ПРОЦЕНУ ДИНАМИЧКЕ
ПОТРОШЊЕ АРИТМЕТИЧКИХ КОЛА
ИМПЛЕМЕНТИРАНИХ НА FPGA

– Докторска дисертација –

Ниш, 2012.

*Пошљи, Господе, утјехе у мраку,
У овој ноћи што се живот зове
Даруј ми духу свијетлу вољу, јаку,
Да увијек ступам на подвиге нове.*

*Славећи Тебе да истину зборим
Другу и брату, силнику и робу;
За правду да се уз праведне борим
И с њима шибам пороке и злобу.*

*Охрабри срце да, у часу сваком,
Вјерује, љуби и опрашта злијем
Кô Твоје сунце нескврњеном зраком –
Да њиме свуда смрзле душе гријем.*

*Пошљи, Господе, утјехе у мраку,
У овој ноћи што се живот зове
Даруј ми духу свијетлу вољу, јаку,
Да увијек ступам на подвиге нове.*

Алекса Шантић, 1908. година

I Аутор	
Име и Презиме	Бојан Јовановић
Датум и место рођења	01.01.1981. Лесковац
Садашње запослење	Универзитет у Нишу Електронски факултет
II Докторска дисертација	
Наслов	Аналитички модел за процену динамичке потрошње аритметичких кола имплементираних на FPGA
Број страница	125
Број слика	67
Број табела	10
Број библиографских јединица	183
Установа и место где је дисертација рађена	Универзитет у Нишу, Електронски факултет
Научна област	Електроника
Ментор	Проф. др Милун Јевтић
III Оцена и одбрана	
Датум пријаве теме дисертације	21. 09. 2011. године
Број одлуке и датум прихватања теме дисертације	НСВ 8/20-01-004/12-016, 29. 05. 2012. године
Комисија за оцену подобности теме дисертације	1. Др Милун Јевтић, ред. проф. 2. Др Ванчо Литовски, ред. проф. 3. Др Бранко Докић, ред. проф. 4. Др Милунка Дамњановић, ред. проф. 5. Др Предраг Петковић, ред. проф.
Комисија за одбрану дисертације	1. Др Милун Јевтић, ред. проф. 2. 3. 4. 5.
Датум одбране	

Научни допринос дисертације:

1. Побољшање тачности модела за процену динамичке потрошње логичке структуре бинарног множача. Побољшање је остварено увођењем новог модела сигнала (ТВТ модела) који, за разлику од до сада коришћеног DBT модела, узима у обзир нелинеарности које се јављају на излазима бинарних множача. Експерименталним путем показана је супериорност новог модела за процену динамичке потрошње. У неким случајевима грешка новог модела (у односу на измерену вредност потрошње) је 4-5 пута мања у односу на DBT модел.
2. Представљање аналитичког модела за процењивање транзиционе активности (тј. динамичке потрошње) логичке структуре бинарног делитеља. Предложени модел је на високом нивоу апстракције система што га чини прилично ефикасним (процене динамичке потрошње добијају се за време реда десетине секунде). Поред тога што омогућава брзу процену динамичке потрошње, експерименталним путем потврђена је и његова висока тачност. Релативне грешке модела за процену динамичке потрошње рачунане у односу на измерене вредности динамичке потрошње не прелазе 15%, што је јако добар резултат имајући у виду чињеницу да је модел на високом нивоу апстракције дизајна.

Листа скраћеница

ARMA	Autoregressive Moving Average
ASIC	Application Specific Integrated Circuit
BJT	Bipolar Junction Transistor
CMOS	Complementary Metal Oxide Semiconductor
DBT	Dual-Bit Type
DSP	Digital Signal Processing
DVS	Dynamic Voltage Scaling
DVTS	Dynamic Vth Scaling
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FPGA	Field Programmable Gata Array
GIDL	Gate Induced Drain Leakage
IIR	Infinite Impulse Response
LNS	Logarithmic Number System
LSB	Least Significant Bit
LUT	Lookup Table
MARWEL	Measurement of Architectural Wire Length
MIPS	Million Instructions Per Second
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MSB	Most Significant Bit
NMOS	Negative-channel Metal Oxide Semiconductor
PAR	Placement And Routing
PMOS	Positive-channel Metal Oxide Semiconductor
PWM	Pulse-Width Modulation
RAM	Random Access Memory
RISC	Reduced Instruction Set Computer
RNS	Residue Number System
SCE	Short Channel Effect
SoC	System-on-Chip
SRAM	Static Random Access Memory
TBT	Triple-Bit Type
TGO	Triple Gate Oxide
VHDL	Very high-speed integrated circuit Hardware Description Language
VLSI	Very Large Scale Integration

Садржај

Листа скраћеница	iv
1 Увод	7
1.1. Мотивација	8
1.1.1. Потрошња електричне енергије	8
1.1.2. Минимизација и оптимизација потрошње	13
1.2. Циљеви дисертације	15
1.3. Организација дисертације	16
2 Технике оптимизације потрошње електричне енергије	19
2.1. Оптимизација фиксне топологије	21
2.1.1. Димензионисање транзистора	22
2.1.2. Подешавање напона напајања - DVS	24
2.1.3. Подешавања напона прага транзистора – DVTS	26
2.1.4. Мулти V_{dd}/V_{th} технике за минимизацију потрошње	27
2.1.5. Гејтовање напона напајања и такта	30
2.1.6. „Гомилање“ транзистора	32
2.1.7. Оптичке интерконекције на чипу	33
2.1.8. Хибридни приступи у минимизацији потрошње	35
2.1.9. CMOS алтернативе за минимизацију потрошње	36
2.1.10. Демонстрација утицаја технолошких параметара на потрошњу	37
2.2. Оптимизација на блоковском нивоу	39
2.2.1. Утицај размештања и повезивања (PAR) на потрошњу	41
2.2.2. Демонстрација PAR утицаја на перформансе система	42
2.2.3. 3D интегрисана кола	43
2.3. Оптимизација на нивоу архитектуре система	44
2.3.1. Методологија селекције оптималне архитектуре	45
2.3.2. Демонстрација методологије селекције оптималне архитектуре	48
2.3.3. Остале минимизационе технике на нивоу архитектуре система	51
3 Метода верификације модела за процену потрошње	55
3.1. Структура FPGA компоненти	56
3.1.1. Virtex II Pro архитектура	57
3.2. Мерна поставка	58
3.2.1. Екстракција капацитивности електричних веза	61
3.2.2. Потрошња логике и улазних бафера	62

3.2.3. XPower алат	63
4 Модели за процену динамичке потрошње логичких структура	65
4.1. Преглед техника за процену потрошње	66
4.2. Моделовање улазног сигнала	67
4.2.1. Сигнали чија је средња вредност једнака 0	68
4.2.2. Сигнали средње вредности различите од 0	72
4.3. Прорачун транзиционе активности	73
4.3.1. Технике прорачуна транзиционе активности	74
4.3.2. Методологија рачунања транзиционе активности	75
4.3.3. Гличеви	79
4.4. ТВТ модел сигнала	80
4.4.1. Евалуација ТВТ модела	81
4.5. Модел за процену потрошње бинарног делитеља	86
4.5.1. Бинарни делитељи	86
4.5.2. Преглед досадашњих активности	88
4.5.3. Рачунање транзиционе активности унутар делитеља	89
4.5.4. Евалуација модела за процену потрошње бинарног делитеља	94
5 Закључци	101
5.1. Минимизација потрошње електричне енергије: закључци	101
5.2. Процена динамичке потрошње електричне енергије: закључци	102
5.3. Будући рад	104
Додаци	105
Литература	113
Апстракт	121
Abstract	123
Биографија	125

Глава 1

Увод

Почетком 1908. године америчком заводу за патенте Ли де Форест (Lee De Forest) пријавио је патент електронске компоненте са три електроде чија је примена била у детекцији електромагнетних таласа. Ова компонента, позната као вакумска цев, триода, била је прва електронска компонента која је омогућавала појачавања електричног сигнала. Претеча савремених електронских компоненти. Де Форест је уствари модификовао већ познату вакуумску диоду тако што је увео нову электроду: решетку. Помоћу напона на овој електроди било је могуће ефикасно контролисати количину електрона који са загрејане катоде доспевају на аноду (анодна струја). Мале промене напона на решетки омогућавале су велику промену анодне струје тј. појачање сигнала. Триоде су нашле своју примену и у првом електронском компјутеру, Ениак-у (ENIAC). Дизајниран 1946. године на Универзитету у Пенсилванији, за потребе америчке војске, овај први рачунар заузимао је површину од преко 60 квадратних метара, имао је потрошњу од 174KW и брзину од 0.05MIPS-а. Тежио је преко 25 тона, садржао 17468 вакуумских цеви (од којих је на сваких сат времена по једна прегоревала те је било потребно замењивати је), 7200 кристалних диода, 1500 релеја (прва електрична имплементација Булове (George Boole) логике), 70000 отпорника, 10000 кондензатора, 5 милиона ручно лемљених спојева. Радио је даноноћно са по шест оператера у свакој смени. У то време откривена је и електронска компонента коју многи сматрају најважнијим открићем XX века (Нобелова награда за физику 1956. године): биполарни транзистор (BJT). Као и триода, имао је три прикључка и омогућавао појачање електричног сигнала. У односу на триоду, био је знатно мањих димензија, поузданији и издржљивији. Такође, био је и мањи потрошач електричне енергије. Шта више, било је могуће смањивати му димензије па је процес минијатуризације могао да почне. Седамдесетих година прошлог века појавила се још једна варијанта транзистора – униполарни MOSFET транзистори. У односу на биполарне имали су своје предности (мањи потрошачи електричне енергије, температурно стабилнији, бржи) и мане (осетљивији на електростатичко наелектрисање, имају мање појачање, могу да издрже мања струјна оптерећења).

Уместо да транзисторе повезују жицама, инжењери из Ферчајлда (Fairchild) и Тексас Инструментса (Texas Instruments) дошли су на идеју да у један комад силицијума ставе неколико транзистора. Родило се прво интегрисано коло (Нобелова награда за физику 2000. године). Захваљујући томе, деценију касније, 1971. године, појавио се први комерцијално доступни 4-битни микропроцесор: Интел (Intel) 4004. На површини од 12mm^2 бројао је 2300 транзистора (ручно дизајниран) и имао исту процесну моћ као и Ениак. Ширина линије везе у овом интегрисаном колу била је $10\mu\text{m}$ (поређења ради, дебљина просечне људске длаке је $100\mu\text{m}$). Од тада па до данас присутан је сталан и незадржив тренд побољшања карактеристика интегрисаних кола. Савремена интегрисана кола производе се уз помоћ нанометарских (nm) технолошких процеса док се електрична енергија коју троше мери микроватима (μW).

1.1. Мотивација

Са усавршавањем технологије производње интегрисаних кола (ширина линије везе код данашњих кола је и до 22nm , што је побољшање од преко 500 пута) било је могуће смањивати димензије транзистора и повећавати степен интеграције. Број транзистора који се имплементирају на једно интегрисано коло дуплирао се на сваке две године [Mo065] па се до данас густина интегрисаних кола повећала преко 65000 пута (данашња интегрисана кола имају на стотине милиона транзистора). Повећан степен интеграције праћен је и већом процесном моћи тј. повећањем брзине рада интегрисаних кола – најновији интелови *i7* процесори обављају инструкције брзином од 159000MIPS-а (да се аутомобилска индустрија развијала сличним темпом данас би било могуће растојање од Ниша до Москве прелазити за 6 секунди). Што се тиче потрошње електричне енергије, са појавом CMOS технологије, веровало се да је проблем потрошње дигиталних кола решен. Статичка потрошња CMOS кола била је изузетно ниска а енергија извора за напајање трошила се само приликом обављања операција тј. приликом промена логичких стања у чворовима интегрисаног кола. Такође, велика паковања раних интегрисаних кола, као и различите технике за дисипацију генерисане топлоте држали су потрошњу кола у границама анvelope дозвољене потрошње. Стога су главни проблеми VLSI дизајнера тога времена били постизање жељених брзина рада и минимизација површине чипа. Кад је број транзистора по јединици површине достигао критичну тачку почели су да се јављају до тада непознати ефекти који су повећавали потрошњу електричне енергије и утицали на то да минимизација потрошње, поред брзине и површине, постане још један критеријум за дизајнере савремених VLSI кола. Да би се оптимизовала потрошња неког дизајна, поред примене минимизационих техника, драгоцено је поседовати и моделе за брзу и тачну процену његове потрошње. Применом ових модела би се у неколико итерација и за кратко време, без потребе за физичком имплементацијом дизајна, могло да дође до његове оптималне архитектуре у погледу потрошње електричне енергије.

1.1.1. Потрошња електричне енергије

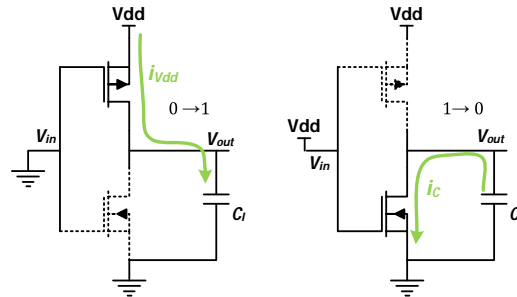
Према начину на који неко интегрисано коло пројектовано у CMOS техници може да троши енергију извора за напајање, потрошњу можемо поделити на:

динамичку P_D , потрошњу кратког споја P_{SC} и статичку потрошњу P_S . Ако се енергија извора троши на успостављање логичких стања у чворовима интегрисаног кола ради се о такозваној динамичкој тј. функционалној потрошњи. Потрошња кратког споја последица је коначног времена трајања растуће односно опадајуће ивице на улазима CMOS кола. У случају статичке потрошње, енергија извора за напајање троши се на неке нежељене процесе који не само да не доприносе функционалности интегрисаног кола већ је и смањују.

Динамичка потрошња

Динамичка потрошња дешава се услед пуњења излазне капацитивности CMOS кола приликом преласка излаза кола из стања логичке '0' у стање логичке '1'. Већа брзина рада неког интегрисаног кола тј. већа радна фреквенција води ка повећању броја транзиција а самим тим и повећању његове динамичке потрошње.

Посматрајмо, на пример, коло инвертора приказано на слици 1.1. Транзистори исцртани испрекиданом линијом су непроводни током одговарајуће транзиције. Сваки пут кад се на улазу инвертора дешава прелаз од логичке '1' на логичку '0' капацитивност излазног чвора инвертора C_l пуни се од 0 до напона V_{dd} . Енергија потребна за ово пуњење обезбеђује се из извора за напајање V_{dd} . Приликом пуњења излазне капацитивности део енергије извора дисипира се на проводном PMOS транзистору.



Слика 1.1: Пуњење и пражњење излазне капацитивности

Енергија коју даје извор за напајање $E_{V_{dd}}$, као и део те енергије који се смешта у излазној капацитивности на крају процеса транзиције може се израчунати интегралом тренутне снаге у границама трајања периода транзиције. На слици 1.2 приказани су таласни облици излазног напона и струје извора током периода пуњења и пражњења излазне капацитивности.

$$E_{V_{dd}} = \int_0^{\infty} i_{V_{dd}}(t) V_{dd} dt = V_{dd} \int_0^{\infty} C_l \frac{dv_{out}}{dt} dt = C_l V_{dd} \int_0^{V_{dd}} dv_{out} = C_l V_{dd}^2 \quad (1.1)$$

$$E_C = \int_0^{\infty} i_{V_{dd}} v_{out} dt = \int_0^{\infty} C_l \frac{dv_{out}}{dt} v_{out} dt = C_l \int_0^{V_{dd}} v_{out} dv_{out} = \frac{C_l V_{dd}^2}{2} \quad (1.2)$$

Из претходних једначина се лако може закључити да се 50% енергије извора троши на пуњење излазне капацитивности чвора а да се друга половина дисипира на проводном PMOS транзистору. Битно је још и приметити да

дисипирана енергија не зависи од величине (а самим тим и од отпорности) PMOS транзистора.

Током транзиције од '0' на '1' на улазу инвертора, C_l се празни преко NMOS транзистора. Током процеса пражњења део акумулиране енергије у C_l дисипира се на овом транзистору.

Иако је енергија коју дисипирају транзистори битна за одређивање такозваних „врућих тачака“ дизајна, електрична енергија коју троши дизајн стриктно се односи на енергију која се обезбеђује преко извора за напајање. Према томе, енергија извора троши се само приликом прелаза 0-1 на излазу инвертора.

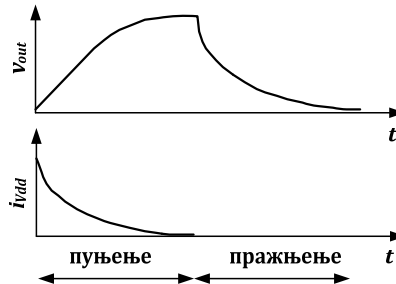
Снага извора за напајање која се троши за пуњење излазне капацитивности током временског интервала T може се израчунати на следећи начин [NM97]:

$$P_D = C_l \cdot V_{dd}^2 \cdot \frac{n_t(T)}{T} \quad (1.3)$$

где $n_t(T)$ представља број 0-1 прелаза током временског интервала T . Према томе, у синхроним колима, просечна потрошња снаге једног гејта рачуна се као:

$$P_D = \alpha \cdot C_l \cdot V_{dd}^2 \cdot f \quad (1.4)$$

где је α просечан број 0-1 прелаза за време трајања једног тактног циклуса, а f фреквенција такта. Будући да свака претходно напуњена капацитивност кад-тад мора да се испразни, број 0-1 прелаза једнак је броју 1-0 прелаза, па се стога параметар α често замењује са $0.5 \cdot sw$, где sw представља транзициону активност тј. просечан број свих прелаза за време трајања једног тактног циклуса. Укупна динамичка потрошња дизајна добија се сабирањем динамичких потрошњи свих гејтова у дизајну.



Слика 1.2: Излазни напон и струја извора током процеса пуњења/пражњења C_l

Посматрајмо, на пример, CMOS чип који ради на фреквенцији од 500MHz и напаја се са 2.5V. Ако претпоставимо да је просечна излазна капацитивност 15fF по гејту, потрошња по гејту биће приближно једнака $50\mu W$. Ако даље претпоставимо да чип садржи милион гејтова (што је реалан случај) и да се прелази дешавају у сваком тактном интервалу ($\alpha=1$) укупна потрошња чипа била би 50W. Ово је, на сву срећу, једна песимистичка прогноза будући да сви гејтови у чипу нису активни у сваком тактном интервалу. Стварна активност у колу углавном је мања ($\alpha < 1$), а прорачун потрошње сложенијих кола често је отежан услед непознавања транзиционе активности. Транзициона активност зависи како од топологије кола тако и од природе и статистике улазних сигнала: ако се улазни сигнали не мењају, нема ни транзиција унутар кола па је динамичка потрошња једнака нули. Са друге

стране, честа транзиција на улазу узрокује повећану транзицију унутар кола па самим тим и повећану динамичку потрошњу.

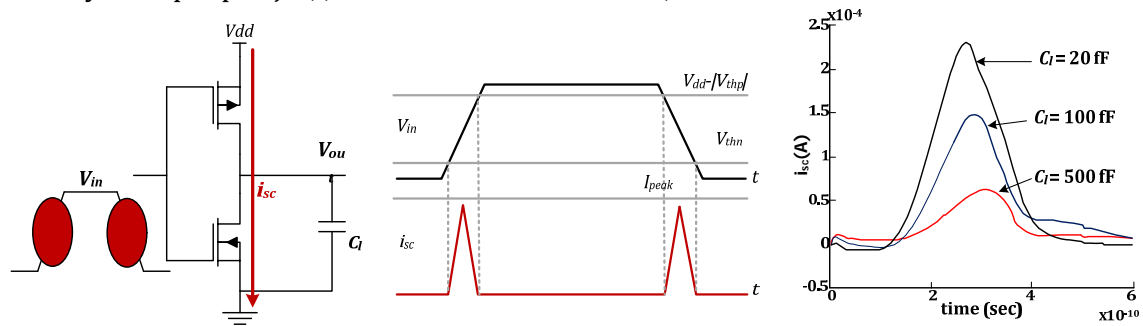
Вредност напона напајања обично је фиксна и константна. Такође, радна фреквенција дизајна је обично унапред позната и дефинисана величина. Према томе, да би потрошња снаге била позната неопходно је одредити два преостала непозната параметра: излазну капацитивност и транзициону активност. Постоји више начина да се нађу ови параметри. О томе ће више бити речи у поглављу које се бави проценом потрошње.

Према својствима ова два непозната параметра динамичка потрошња се даље може поделити на три компоненте:

- Потрошња дистрибуционе мреже такта са транзиционом активношћу која је константна и једнака 2 (за време трајања тактног интервала имамо један 0-1 и један 1-0 прелаз).
- Потрошња функционалних јединица и меморија дизајна тј. потрошња логике. Параметар C_i у овом случају одговара излазној капацитивности логичких гејтова.
- Потрошња интерконекција (веза) којима су функционалне јединице повезане. C_i је у овом случају одређена типом и дужином електричне везе која се користи за повезивање.

Потрошња кратког споја

Коначно време трајања растуће односно опадајуће ивице на улазу CMOS инвертора за последицу има стварање проводног пута између извора за напајање и масе, тј. протицање струје кратког споја. Са слике 1.3 може се видети да ова струја протиче за време док је напон на улазу инвертора већи од напона прага NMOS транзистора а мањи од напона напајања умањеног за напон прага PMOS транзистора. Максимум ове струје зависи од струја засићења MOS транзистора, а достиже се кад је улазни напон приближно једнак половини напона напајања. Максимум струје кратког споја зависи и од односа трајања ивице на улазу и излазу инвертора тј. од величине излазне капацитивности C_i .



Слика 1.3: Струја кратког споја на излазу CMOS инвертора

У случају веома велике излазне капацитивности време трајања опадајуће ивице на излазу инвертора много је веће од времена трајања растуће ивице на његовом улазу. Прелаз са '0' на '1' на улазу се деси толико брзо да PMOS транзистор не стигне да постане проводан па је струја кратког споја приближно једнака нули. У

супротно, када је излазна капацитивност веома мала време трајања опадајуће ивице на излазу инвертора много је мање од времена трајања растуће ивице на његовом улазу. Напон између сорса и дрејна PMOS транзистора током већег дела трајања транзиционог периода једнак је напону напајања V_{dd} што за последицу има максималну струју кратког споја (једнаку струји засићења PMOS транзистора). Из израза за струју кратког споја:

$$I_{sc} = K \cdot (V_{dd} - 2V_{th})^3 \cdot \tau \cdot f \quad (1.5)$$

може се видети да она поред трајања растуће/опадајуће ивице τ , зависи још и од технолошких параметара представљених константом K , напона напајања и прага, као и од брзине рада кола, f [BG05].

Многи аутори бавили су се проблемом моделовања енергије која се дисипира услед протицања струје кратког споја [Vee84, VS94, WV98, JBK98, AAN03, Bis10]. Модели су мање или више тачни будући да у обзир узимају мање или више параметара и ефеката који се јављају у CMOS колима. У [NS00a] аутори, поред тога што се баве проценом потрошње кратког споја, покушавају да предвиде шта ће бити са овом потрошњом у CMOS колима будућих генерација. Имајући у виду очигледан тренд смањивања напона напајања V_{dd} и напона прага транзистора V_{th} аутори упозоравају да би, у колико се не задржи данашњи однос ове две величине, тј. ако се напон прага брже смањује од напона напајања (смањивањем напона прага транзистори постају бржи па се може рећи да је потреба за смањивањем напона прага реална), потрошња кратког споја могла да порасте и да чини значајан део (и до 20%) укупне потрошње.

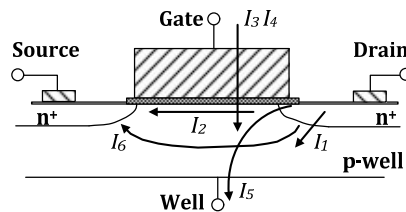
Једна од веома ефикасних техника за минимизацију потрошње кратког споја је техника усаглашавања трајања растућих/опадајућих ивица улазних и излазних сигнала. На тај начин обезбеђује се да потрошња кратког споја буде занемарљива и да чини мање од 10% од укупне потрошње. Стога многи аутори у анализама потрошње електричне енергије потрошњу кратког споја често не узимају у обзир. Треба рећи и да се смањивањем напона најчапача потрошња кратког споја такође смањује. У изузетном случају, кад би напон напајања био мањи од збира напона прагова N-каналног и P-каналног транзистора тј. $V_{dd} < V_{thn} + |V_{thp}|$, потрошња кратког споја била би потпуно елиминисана будући да оба транзистора никад не би била проводна у исто време [RCN03].

Статичка потрошња

Различите струје цурења које се углавном јављају за време док је транзистор непроводан и када би, идеално, требало да се понаша као бесконачна отпорност узрок су статичке потрошње. Управо су мање струје цурења у непроводном стању биле један од главних разлога за прелазак са биполарних на CMOS транзисторе (крајем '80-их и почетком '90-их година прошлог века). Тада се веровало да је проблем статичке потрошње решен. Како су се, међутим, димензије транзистора последњих деценија константно смањивале [RMM03], предности које су се добијале услед мањих димензија (повећање густине и перформанси кола, као и смањивање потрошње) полако су почеле да нестају (почев од 100 nm) услед квантних ефеката који су главни узрок повећања цурећих струја па самим тим и статичке потрошње. Будући да постаје све значајнији део укупне потрошње,

статичкој потрошњи потребно је посветити пажњу приликом дизајнирања савремених VLSI кола. Тачне и ефикасне процене ове потрошње током фазе дизајнирања система драгоцене су јер дају увид у потрошњу дизајна који ће се имплементирати, а самим тим и одговор на питање да ли су испуњене спецификације у погледу потрошње. Прецизни симулатори, као HSPICE™ на пример, дају врло добре процене цурећих струја, али им је недостатак што су применљиви само на мања кола (због проблема конвергенције, процесорског времена и меморије). Модели који се баве физиком механизма цурећих струја [SSK+87, AMK+05] исувише су компликовани да би били примењивани од стране дизајнера. Са друге стране, неки од бржих модела за процену ових струја предложени су [CJL+98, CLL+00, MRR05].

Две најдоминантије струје цурења су (види слику 1.4): струја цурења у каналу (I_2) и струја тунеловања оксида гејта (I_3). Као последица агресивног смањивања димензија транзистора још неке цуреће струје добијају на значају последњих година.



Слика 1.4: Струје цурења у MOS структури [RMM03]

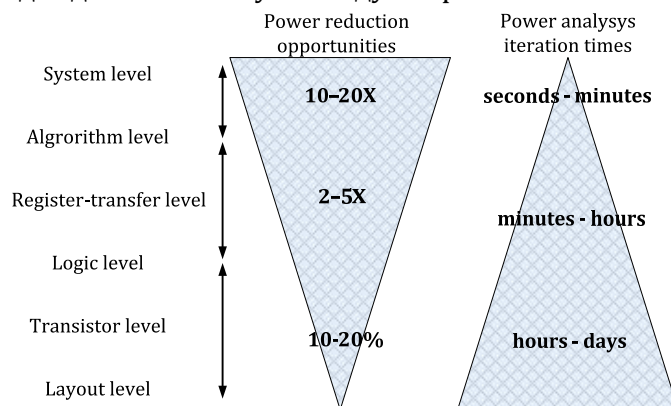
Оне се јављају услед смањивања дужине канала: струја ињекције „врхних“ носилаца из субстрата у оксид гејта (I_4) као и струја пробоја канала (I_6), услед смањивања дебљине оксида гејта: гејтом индукована струја дрејна – GIDL (Gate Induced Drain Leakage) (I_5) и услед повећане концентрације донора у областима сорса и дрејна: струја инверзне поларизације pn споја (I_1). У поглављу које се бави техникама за минимизацију статичке потрошње механизмима ових струја биће детаљније објашњени.

Треба рећи да су цуреће струје, нарочито струја цурења у каналу као најдоминантнија међу њима, јако зависне од температуре [AMK+05]. Како се вредност ових струја експоненцијално повећава са порастом температуре, на 85°C (што је уобичајени лимит температуре p - n споја за комерцијалне електронске компоненте) ове струје су и до 60 пута веће него на собној температури. Тако, неки електронски уређај који на собној температури ради исправно, на повећаној температури може да троши знатно више енергије и да при том има отказе у раду.

1.1.2. Минимизација и оптимизација потрошње

Диктирана захтевима за повећањем функционалности и перформанси, и у исто време смањивањем цене, минијатуризација тј. скалирање силицијумске технологије одвија се преко четрдесет година. Као последица тога, повећани степен интеграције повећао је и потрошњу чипова. Одвођење топлоте и прерасподела снаге једни су од главних проблема са којима се сусрећу дизајнери данашњих интегрисаних кола.

Последњих година, једна од битних особина многих електронских уређаја постала је и портабилност (преносивост). Иако се, историјски гледано, од портабилних уређаја захтевала мала процесна моћ (дигитални ручни сат нпр.), данас то није случај. Међу новим портабилним уређајима су бројни ручни мултимедијални уређаји са могућностима видео и звучног записа, препознавања гласа, рукописа итд. Велика процесна моћ ових уређаја треба да буде праћена и њиховом ниском потрошњом будући да им је напајање батеријско. Стога се потрошњом електричне енергије треба позабавити у сваком од нивоа пројектовања VLSI интегрисаних кола приказаним на слици 1.5. Процес пројектовања састоји се у томе да се у свакој наредној фази пројектовања систем представља са мало више детаља него у претходној фази. Почиње се од највишег нивоа апстракције (ниво система) у коме се дизајн представља скупом апстрактних комуникационих процеса или задатака, без знања о томе да ли ће се ови задаци имплементирати у хардверу или компајлирати у софтверу неког *embedded* процесора. На крају се, у нивоу лејаута, завршава физичким смештањем и повезивањем стандардних ћелија на чипу, одређивањем технолошких параметара (димензионисање транзистора и електричних веза за повезивање) и генерисањем мреже такта и напајања. Тек након ове фазе могуће је екстраковати маске помоћу којих се врши производња чипова. Да би се добио дизајн оптималне потрошње електричне енергије неопходно је у сваком од нивоа пројектовања примењивати технике минимизације и естимације (процене) потрошње. Под минимизацијом потрошње подразумева се примена разних техника које доводе до дизајна са минималном потрошњом, под условом да се при том не наруше његове, унапред задате, спецификације. Током процене потрошње прорачунава се потрошња дизајна (са одређеном тачношћу) и тиме на неки начин повећава поузданост пројектовања будући да добијамо информацију о томе да ли су испуњени унапред задати захтеви у погледу потрошње.



Слика 1.5: Могућности уштеде енергије и време трајања процене потрошње на различитим нивоима пројектовања дизајна [Rab09]

На слици 1.5 приказане су и могућности уштеде енергије као и времена потребна за процену потрошње на различитим нивоима пројектовања. Као што се са слике може видети, највећу уштеду потрошње (10 до 20 пута) могуће је остварити у почетним фазама пројектовања, на највишим нивоима апстракције дизајна. У исто време, процена потрошње на овим нивоима није временски захтевна што омогућава да се за кратко време испита велики број различитих архитектура дизајна – кандидата за најмању потрошњу. Утицати на потрошњу а при том не жртвовати перформансе система постаје све теже и теже на нижим

нивоима пројектовања. Тако је на нивоу лејаута могуће остварити уштеду у потрошњи од 10 до 20%, док је за информације о потрошњи дизајна потребно чекати сатима, па чак и данима будући да је дизајн представљен са највише детаља па је потребно обрадити веома велику количину података [RJD98, Kri04].

1.2. Циљеви дисертације

Главни циљеви ове дисертације су преглед техника које се, на различитим нивоима пројектовања система, користе за минимизацију статичке и динамичке потрошње електричне енергије, развој брзих и флексибилних модела (на високом нивоу апстракције) за процену динамичке потрошње логике DSP (Digital Signal Processing) кола имплементираних на FPGA, као и верификација предложених модела кроз поређење са измереном потрошњом. Могу се сумирати на следећи начин:

- Механизми и типови потрошње електричне енергије у савременим CMOS колима.
- Преглед техника за минимизацију статичке и динамичке потрошње електричне енергије:
 - а) Минимизационе технике технолошког нивоа. Параметри преко којих дизајнер утиче на потрошњу електричне енергије су технолошки – димензије и типови транзистора, бирање напона прага транзистора, напона напајања итд. Демонстрација утицаја технолошких параметара на укупну потрошњу кроз имплементацију референтног дизајна бинарног делитеља на FPGA компоненте различитих технолошких карактеристика тј. фамилија (Virtex-4, Virtex-5, Virtex-6, Virtex-6 Lower Power).
 - б) Минимизационе технике блоковског нивоа – различити начини физичке имплементације дизајна, 3Д интегрисана кола итд. Демонстрација утицаја физичке имплементације на перформансе система кроз имплементацију референтног дизајна бинарног делитеља на Xilinx FPGA компоненту из Virtex-5 фамилије.
 - в) Минимизационе технике на нивоу архитектуре – паралелизам и проточност (pipelining) различитих нивоа, RNS (Residue Number System) и LNS (Logarithmic Number System) аритметика итд. Демонстрација утицаја архитектуре на потрошњу кроз имплементацију референтног дизајна на ASIC (Application Specific Integrated Circuit) чип (у технологији AMIS 0.35) и Xilinx FPGA компоненте из Virtex фамилије.
- Анализа најпогоднијег модела за моделовање улазних сигнала DSP блокова који омогућава брзо и лако прорачунавање транзиционе активности:
 - а) DBT (Dual Bit Type) модел сигнала.
 - б) TBT (Triple Bit Type) модел сигнала за моделовање нелинеарности које се јављају на излазу логичке структуре бинарног множача а коју DBT модел сигнала не узима у обзир.
- Опис методологије за прорачунавање транзиционе активности

- Развој модела (високог нивоа) за процену динамичке потрошње логичких структура DSP кола имплементираних на FPGA што подразумева моделовање излазне капацитивности DSP модула имплементираних на FPGA на високим нивоима апстракције (док је дизајн још увек на системском нивоу или на нивоу алгоритма).
- Верификација предложених модела кроз поређење са измереном потрошњом што подразумева:
 - а) Опис мерног система који омогућава тачно мерење потрошње као и раздвајање различитих компоненти потрошње (статичку потрошњу, потрошњу такта и потрошњу логике и интерконеција).
 - б) Показивање валидности модела за различиту статистику и различит број битова улазног сигнала.

1.3. Организација дисертације

Будући да ова дисертација обрађује неколико различитих области поменутих у претходном поглављу, аутор ће се трудити да, где год је то могуће, да преглед активности и резултата који су до сада постигнути, као и да изврши поређење тих резултата са резултатима добијеним током рада на дисертацији. Организација дисертације приказана је на слици 1.6.



Слика 1.6: Организација дисертације

Глава 2 бави се техникама за оптимизацију тј. минимизацију укупне потрошње електричне енергије. Дат је преглед минимизационих техника које се примењују на нивоу фиксне топологије (када се на потрошњу утиче преко технолошких параметара дизајна као што су димензије транзистора, напон напајања и напон прага), на блоковском нивоу (на потрошњу се утиче преко различитих техника физичке имплементације дизајна) и на нивоу архитектуре

дизајна (на потрошњу се утиче преко различитих трансформација архитектуре дизајна).

Главе 3 и 4 баве се техникама за процену динамичке потрошње. Најпре је у глави 3 описана метода верификације тј. провере тачности модела за процену потрошње. Представљена је мерна поставка и методологија екстракције динамичке потрошње логике из вредности укупне измерене потрошње.

У глави 4 описана је методологија за процену динамичке потрошње логике високог нивоа, базирана на логичкој структури компоненте и аналитичком прорачуну транзиционе активности која се јавља унутар компоненте. Такође је описана и примена ове методологије за процену динамичке потрошње *ripple carry array* бинарних множача и различитих типова *radix-2 non-restoring* бинарних делитеља, као и резултати њихове евалуација.

На крају, глава 5 резервисана је за закључке дисертације и предлоге за будући рад.

Глава 2

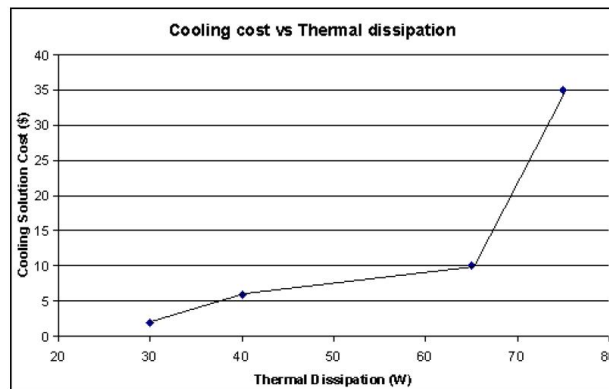
Технике оптимизације потрошње електричне енергије

КАКО потрошња електричне енергије полако постаје фактор који ограничава перформансе савремених интегрисаних кола, ова глава посвећена је техникама које за циљ имају постизање минимизације тј. оптимизације потрошње. Примене ових оптимизационих техника део су процеса пројектовања дизајна „свесног потрошње“ (power-aware design process). Најпре је у уводу главе кроз неколико илустративних примера демонстрирана реалност потребе за минимизацијом потрошње. Такође су приказани и дугорочни трендови у погледу потрошње интегрисаних кола. Указано је и на разлику између минималне и оптималне потрошње неког интегрисаног кола. Након тога дат је детаљан преглед техника за минимизацију потрошње које се примењују на различитим нивоима пројектовања интегрисаног кола. На сваком од нивоа је кроз пример дизајна бинарног делитеља демонстриран утицај различитих параметара на потрошњу електричне енергије.

MP3 плејери, мобилни телефони, PDA (Personal Digital Assistant) и GPS (Global Positioning System) уређаји, таблет и лаптоп рачунари захтевају велику процесну моћ и комплексну функционалност уз истовремену малу потрошњу електричне енергије. У случају да су састављени од компоненти које нису оптимизоване у погледу потрошње електричне енергије, ови портабилни уређаји конзумирају око 40W енергије [Tod06]. Кад би овакви уређаји између два пуњења функционисали око 10 сати, са Литијум-Јонским батеријама енергетске густине 100Wh/kg, тежина батерије требала би да буде 4kg. У случају лаптоп рачунара који процесира неку мултимедијалну апликацију трошећи при том 24.8W енергије [KCN+01], при истом сценарију као у претходном примеру (10 сати рада): батерија лаптопа би требало да тежи 2.48kg. Ова два примера потврђују да би без примене техника за минимизацију потрошње електричне енергије, портабилни батеријски напајани уређаји имали озбиљне функционалне проблеме.

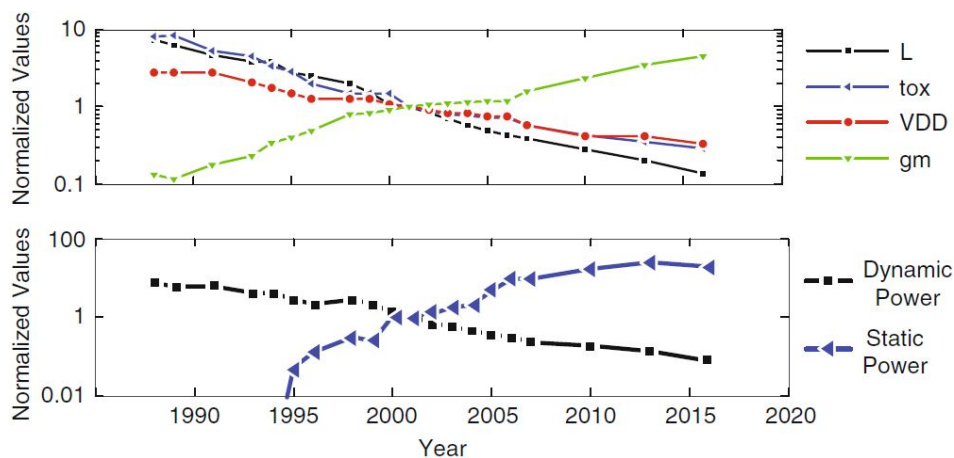
Што се тиче потрошачке електронике, уштеда енергије значи и уштеду новца. Што је мања дисипације енергије електронских система мање је и загревање просторије у којој је систем смештен, мања потрошња струје па према томе и мањи утицај на глобалну околинду, мања потреба за одвођењем топлоте из просторије и мање је буке у њој (елиминација вентилатора за хлађење). Такође, ако се минимизацији потрошње интегрисаног кола не посвети дужна пажња више новца трошиће се на скупа кућишта и разне технике расхлађивања чипа. Илустрација ове тврдње приказана је на слици 2.1 [GBC+01]. Приликом пораста потрошње електричне енергије тј. термалне дисипације зависност између дисипиране топлоте и цене техника за хлађење је нелинеарна. Мало повећање дисипиране топлоте за последицу има велике додатне трошкове у систему за хлађење.

Што се трендова у погледу димензија транзистора и потрошње електричне енергије тиче, на слици 2.2 приказане су неке пројекције у развоју полупроводничке електронике учињене још 2001. године [IRTS]. Сви параметри на графиконима нормализовани су у односу на њихове номиналне вредности из 2001. године.



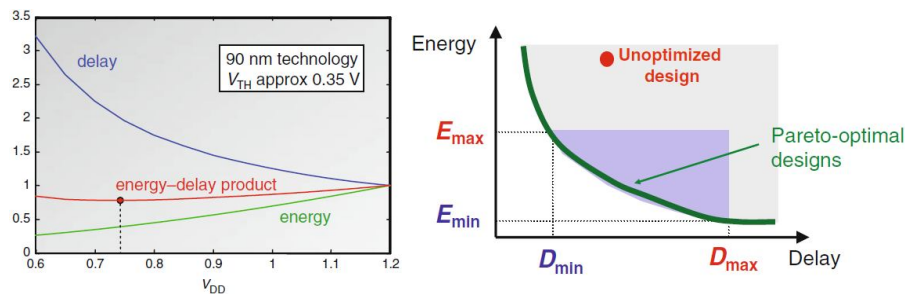
Слика 2.1: Трошкови хлађења система у функцији термалне дисипације

Са слике се може видети да се смањивање дужине канала (L) одвија агресивније у односу на смањивање дебљине оксида гејта и напона напајања. Такође, приметно је рапидно повећања статичке потрошње електричне енергије.



Слика 2.2: IRTS предвиђања за димензије транзистора и потрошњу електричне енергије

Из свега претходно наведеног може се закључити да је минимизацији потрошње електричне енергије, како статичке тако и динамичке, неопходно посветити дужну пажњу. Проблем оптимизације потрошње је вишедимензионалан, што значи да се до дизајна са оптималном потрошњом електричне енергије може доћи једино онда када се о потрошњи води рачуна у свакој фази дизајнирања система (види слику 1.5). Такође, оптимизационе технике морају међусобно бити избалансиране како би минимизациони потенцијали сваког од параметара који утиче на потрошњу били максимално искоришћени. Треба имати на уму и то да оптимално не значи увек и минимално зато што су минимална потрошња електричне енергије и минимално кашњење дизајна међусобно опречни захтеви (види слику 2.3). Тако се може десити да обраћањем пажње само на минимизацију потрошње добијемо дизајн са неприхватљивим кашњењем. Потребно је дакле правити компромисе, па коректна оптимизација типично или минимизира потрошњу енергије за унапред задата ограничења у погледу брзине рада или максимизира процесну моћ за унапред задату дозвољену потрошњу. За неки дизајн и скуп његових параметара могуће је извести парето-оптималну криву која за било коју вредност кашњења дизајна даје минималну могућу потрошњу и обрнуто (десни график на слици 2.3) [MSN+04]. Као мера квалитета дизајна често се уводи још један параметар – производ између енергије и кашњења (EDP – Energy Delay Product).



Слика 2.3: Однос потрошње и кашњења у интегрисаном колу

Грубо гледано, три су нивоа на којима је могуће утицати на потрошњу електричне енергије неког дизајна: ниво оптимизације архитектуре (спољашња петља), оптимизација на блоковском нивоу (средња петља) и оптимизација фиксне топологије (унутрашња петља) [BHM+02]. О оптимизационим техникама које се примењују на сваком од ових нивоа биће речи у наставку ове главе.

2.1. Оптимизација фиксне топологије

Оптимизационе технике на овом нивоу не мењају топологију кола већ преко технолошких параметара као што су димензије транзистора, напон напајања и напон прага покушавају да оптимизују потрошњу електричне енергије. Неки истраживачи испитују утицај појединачног параметра на потрошњу и брзину дизајна док други обављају темељнију анализу испитујући међусобни утицај два или више параметара. Већина техника за минимизацију потрошње електричне енергије на овом нивоу примењује се на некритичне путање у дизајну будући да на тим местима постоји временска „лабавост“. Применом минимизационе технике транзистори у некритичној путањи постају мањи потрошачи по цену смањивања брзине рада. Тиме се не нарушавају унапред задата ограничења у погледу брзине

рада дизајна, а ипак му се минимизира потрошња. Неке општеприхваћене и широко коришћене технике за минимизацију потрошње на овом нивоу су: димензионисање транзистора, подешавање напона напајања/прага, мулти V_{dd} и мулти V_{th} технике, гејтовање напона напајања и такта, „гомилање“ транзистора, увођење слоја оптичке интерконеције, употреба алтернативних (нано) компоненти итд.

2.1.1. Димензионисање транзистора

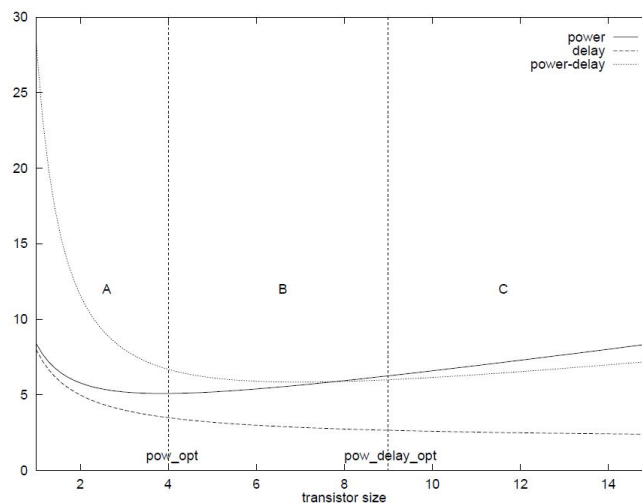
У почетку је димензионисање транзистора коришћено као веома ефикасна техника за смањивање површине интегрисаних кола у функцији задатих временских ограничења или за налажење компромиса између површине и кашњења кола [FD85, СК88, КО192, SRV+93, СЕМ99]. Алгоритми за оптимизацију димензија транзистора били су углавном брзи и итеративни (TILOS, EinsTuner итд.) и водили су до оптималних димензија транзистора. Потрошња није била у првом плану а аутори су наглашавали да је потрошња пропорционална „активној површини“ и да се минимизацијом површине углавном (мада не увек) минимизира и потрошња. Са побољшањем технолошког процеса и сталним трендом смањивања димензија транзистора, критеријум минималне површине временом је замењен критеријумом минималне потрошње електричне енергије. Димензије транзистора које се најчешће мењају су ширина канала, W и дебљина оксида, t_{ox} . Димензионисањем ширине канала смањује се капацитивност кола и тиме минимизира динамичка потрошња, док димензионисање (повећање) дебљине оксида гејта минимизира струје цурења оксида гејта тј. статичку потрошњу.

За транзисторе са дугим каналом (већим од $1\mu\text{m}$) минималне димензије транзистора гарантовале су и минималну потрошњу. Ово правило, међутим, не важи са под-микронске CMOS транзисторе. Ефекти кратког канала, пре свега засићење брзине носилаца и повећане паразитне капацитивности, утицали су на то да оптималне димензије транзисторе не буду више оне минималне. У анализи ограниченој на просте логичке гејтове и ланце инветора [RK97, RCN03] су на бази аналитичког прорачуна потрошње CMOS кола показали да повећање димензија транзистора омогућава додатно смањивање напона напајања а тиме и додатну уштеду потрошње. Ово повећање димензија транзистора корисно је само до одређене границе. Предимензионисање транзистора тј. повећање димензије преко поменуте границе негативно ће утицати на перформансе кола што би као компензацију захтевало повећање напона напајања тј. потрошње дизајна. Под димензијама транзистора аутори подразумевају ширину његовог канала, W . Анализирајући ланац од два CMOS инвертора [RK97] извештавају да смањивање напона напајања омогућено налажењем оптималне ширине канала транзистора омогућава уштеду у потрошњи и до 50%. У [BOI95] аутори дају исцрпнију анализу утицаја ширине канала на потрошњу и кашњење CMOS кола. На бази аналитичког модела за потрошњу CMOS кола који узима у обзир и потрошњу кратког споја аутори дају алгоритме за проналажење оптималних ширина канала (W_n , W_p) транзистора које гарантују минималну потрошњу као и за налажење оптималних W_n , W_p које минимизирају потрошњу али за унапред задата ограничења у погледу брзине рада кола. У исто време, баве се и анализом чинилаца који утичу на димензионисање транзистора (време трајања растуће/оппадајуће ивице сигнала

на улазу кола, fanout, величина оперећења кола итд.). Резултати рада сумирани су на слици 2.4 која приказује зависности потрошње (*power*), кашњења (*delay*) и њиховог производа (*power-delay*) од димензија CMOS инвертора са fanout-ом једнаким 5. Графици су цртани на бази једначина које су изведене у раду. *pow_opt* и *pow_delay_opt* су оптималне димензије транзистора које гарантују минималну потрошњу и минималну потрошњу за унапред задата временска ограничења, респективно. Када су димензије транзистора мање од *pow_opt* и потрошња и кашњење кола смањују се са повећањем димензија. Као резултат тога производ потрошње и кашњења опада врло брзо (регион А на слици 2.4). Када су димензије транзистора у границама (*pow_opt*, *pow_delay_opt*), са даљим порастом димензија потрошња се повећава док се кашњење транзистора још увек смањује (регион В). Производ потрошње и кашњења идаље опада, али знатно спорије него у региону А. На крају, када су димензије транзистора веће од *pow_delay_opt*, производ потрошње и кашњења почиње да расте услед тога што повећање потрошње доминира у односу на смањење кашњења (регион С).

Према томе, ако се транзистор налази у региону А повећањем његових димензија могуће је у исто време смањити потрошњу и кашњење кола. Свим транзисторима требало би повећати димензије, барем до *pow_opt*. У региону В смањивање кашњења доминира у односу на повећање потрошње па је према томе и овде уносно повећавати димензије транзистора. Када је транзистор у региону С, нагиб криве производа потрошње и кашњења одређује ефективност димензионисања транзистора. Будући да овај нагиб одређује компромис између потрошње и кашњења, од два транзистора у региону С онеме са мањим нагибом криве ефективније је повећавати димензије.

Треба, међутим, нагласити и то да је углавном немогуће (тј. Дизајнерима нија на располагању) континуално мењање димензија транзистора. У већини случајева дизајнери могу да баратају са две или више дискретних вредности ширине канала



Слика 2.4: Зависност потрошње, кашњења и њиховог производа од димензије CMOS инвертора

транзистора тј. дебљине оксида гејта које су им на располагању из библиотека стандардних ћелија. У [НОК01] дато је правило по коме се рачунају дискретне вредности ширина канала транзистора. На слици 2.5 ово правило је дато за случај

две (W_1, W_2), три (W_1, W_2, W_3) или четири (W_1, W_2, W_3, W_4) дискретне вредности за ширину канала транзистора.

3a $\{W_1, W_2\}$:	$W_2 = 1/2W_1$
3a $\{W_1, W_2, W_3\}$:	$W_2 = 2/3W_1, W_3 = 1/3W_1$
3a $\{W_1, W_2, W_3, W_4\}$:	$W_2 = 3/4W_1, W_3 = 1/2W_1, W_4 = 1/4W_1$

Слика 2.5: Правило за рачунање дискретних вредности ширина канала транзистора [НОК01]

Аутори такође наглашавају да не постоји разлог због којег би требало користити више од 3 дискретне вредности за ширине канала транзистора. Употребом транзистора са различитим ширинама канала делује се на транзисторе у некритичној путањи. Овим транзисторима смањује се ширина канала што за последицу има минимизацију њихове потрошње уз истовремено повећање њиховог кашњења. Будући да транзистори нису у критичној путањи дизајна целокупне перформансе система се не смањују. Применом ове технике (уз употребу транзистора са три различите ширине канала) на референтни дизајн измерена је уштеда у потрошњи од 15% [НОК01].

На сличан начин могуће је вршити минимизацију потрошње (углавном статичке) помоћу транзистора са две или више дискретних вредности за дебљину оксида гејта. Смањивање дебљине оксида гејта негативно се одражава на статичку потрошњу (нарочито почев од 100nm технологије па надаље). С друге стране, сталан тренд смањивања димензија канала транзистора условљава смањивање дебљине оксида гејта како би се одржала његова струјна моћ. Транзистор са тањим оксидом гејта је већи потрошач електричне енергије (има веће цуреће струје) али је и бржи од транзистора са дебљим оксидом гејта. Употреба транзистора са дебљим оксидом гејта у некритичној путањи неће негативно утицати на перформансе кола док ће са друге стране минимизирати његову потрошњу. У [LDB+04] аутори су користили транзисторе са две различите дебљине оксида гејта са циљем да минимизирају статичку потрошњу. Техника је примењена на сет benchmark дизајна, а струје цурења смањене су и до 5 пута што је плаћено повећањем кашњења кола од свега 5%. На сличан начин могуће је користити и транзисторе са три различите дебљине оксида гејта (triple gate oxide – TGO). Трећи оксид углавном се користи за улазно/излазне и мешовите (аналогно-дигиталне) сигнале [Mac04].

Употреба high-k материјала са великом диелектричном константном такође је једно од ефикасних решења за минимизацију струја цурења гејта. Овакви материјали омогућавају и додатно смањивање дебљине оксида.

2.1.2. Подешавање напона напајања - DVS

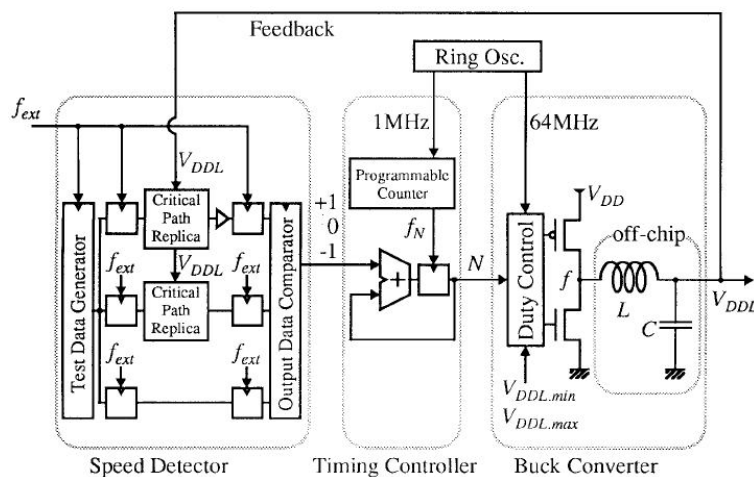
Техника подешавања напона напајања (DVS – Dynamic Voltage Scaling) састоји се у томе да се у сваком тренутку систем напаја оптималном (тј. минималном могућом) вредношћу напона напајања под условом да се не нарушавају унапред задата ограничења у погледу његове брзине рада. Ова техника први пут је представљена 1998 године [KSM+98]. Након тога је доста разрађивана и примењивана [НКQ+99, BPS+00, PLS01, JG04, CC05]. Шема за одређивање оптималне вредности напона напајања приказана је на слици 2.6. Напаја се из

спољашњег извора за напајање V_{DD} а на свом излазу даје оптималну вредност за напон напајања (V_{DDL}) са којим систем може да функционише на задатој радној фреквенцији f_{ext} . Састоји се из три дела: 1) Виск конвертора, 2) временског контролера и 3) детектора брзине. Виск конвертор на свом излазу генерише $(N/64) \cdot V_{DD}$ за интерни напон напајања V_{DDL} . N је цео број из опсега од 0 до 63 и добија се са излаза временског контролера. Према томе, за вредност екстерног напона напајања од $V_{DD}=3.3V$ резолуција са којом је могуће мењати V_{DDL} је 50mV. *Duty Control* коло генерише правоугаоне таласне облике са фактором испуне (duty cycle) $N/64$. Просечна вредност напона ових импулса одређена је нископропусним филтром другог реда који се екстерно конфигурише помоћу калема, L и кондензатора, C . *Duty Control* колу такође је могуће подесити минималну и максималну вредност за интерни напон напајања, V_{DDLmin} и V_{DDLmax} .

Временски контролер прорачунава вредност N уз помоћ 10-битног програмабилног бројача и излаза из детектора брзине. Вредност +1 на излазу детектора значи да се променљива N а самим тим и интерни напон најпајања V_{DDL} мора повећати док -1 значи да се N (V_{DDL}) смањује.

Детектор брзине надгледа кашњење у критичној путањи у систему, и то уз помоћ њених реплика које се напајају са V_{DDL} . Кад је интерни напон напајања V_{DDL} исувише мали да би систем напајан њиме могао да функционише на радној фреквенцији f_{ext} детектор брзине на свом излазу даје +1 како би се вредност V_{DDL} повећала. С друге стране, када је вредност интерног напона напајања исувише велика, детектор брзине на свом излазу генерише -1 што за последицу има смањивање V_{DDL} . Уз помоћ овакве контроле помоћу повратне спреге (feedback) могуће је одредити минималну вредност V_{DDL} на којој ће систем бити у стању да функционише на радној фреквенцији f_{ext} . Из безбедносних разлога реплици критичне путање додато је мало, додатно кашњење.

Будући да је циклус одређивања брзине (базиран на f_{ext}) много мањи у односу на временску константу нископропусног филтра контрола уз помоћ повратне спреге би могла да проосцилује. Због тога, да би се обезбедио брз и поуздан одговор контроле повратне петље, програмабилни бројач функционише са акумулационом фреквенцијом f_N . Што је већа фреквенција f_N бржи је одговор петље повратна спрега а стабилност је мања.



Слика 2.6: Шема за одређивање оптималне вредности за напон напајања

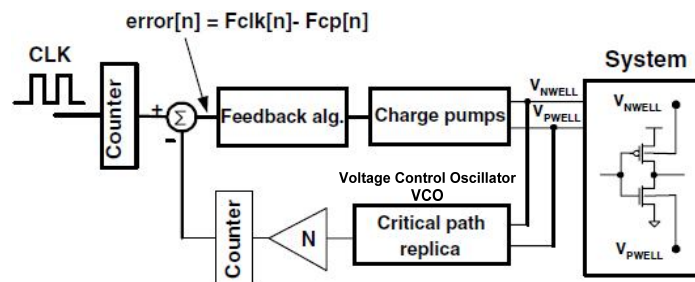
Ову технику заједно са техником подешавања напона прага о којој ће бити речи у наставку аутори у [KSM+98] применили су на 32-битни RISC процесор R3900 пројектован у 0.4 μ m CMOS технологији. Шеме за одређивање оптималних вредности напона напајања и напона прага незнатно (мање од 1%) су повећале површину чипа док је уштеда енергије мерена у MIPS/W била више него дупла.

2.1.3. Подешавања напона прага транзистора – DVTS

Имајући у виду да је динамичка потрошња сразмерна квадрату напона напајања смањивање напона напајања вероватно најефикасније минимизира динамичку потрошњу. Међутим, са смањивањем напона напајања транзистори у колу постају спорији што за последицу има смањивање перформанси целокупног система. Овај проблем може се решити смањивањем напона прага (V_{th}) транзистора. Транзистори сада имају бржи одзив али је невоља што се са смањивањем напона прага струје цурења експоненцијално повећавају. Може се дакле закључити да је напон прага транзистора још један од параметара преко кога је могуће контролисати статичку потрошњу електричне енергије.

Код транзисторске структуре са задњим гејтом динамичка контрола напона прага врши се помоћу напона на задњем гејту [YVC+97, Sri00]. У стандардној CMOS транзисторској структури напон прага могуће је контролисати помоћу напона (V_{BS}) на четвртом изводу – субстрату. Електрична шема (DVTS – Dynamic V_{th} Scaling) која у сваком тренутку рада система помоћу напона на субстрату одређује оптималну вредност (са становишта минималне потрошње) напона прага транзистора приказана је на слици 2.7 [KR02].

DVTS шема континуално одређује напоне на субстратима PMOS и NMOS транзистора са циљем да се минимизира статичка потрошња система а не наруше претходно задата временска ограничења. Најнижа вредност за напон прага транзистора постиже се када субстрат није поларисан ($V_{BS}=0V$). У том случају перформансе система али и статичка потрошња су максималне. Инверзним поларисањем субстрата ($V_{BS}>0$ за NMOS, $V_{BS}<0$ за PMOS) напон прага транзистора се повећава што утиче на смањивање перформанси и статичке потрошње система. Поларисање субстрата врши се помоћу струјне пумпе (charge pump). Напоном контролисани осцилатор (Voltage-Controlled Oscillator – VCO) конвертује напон субстрата



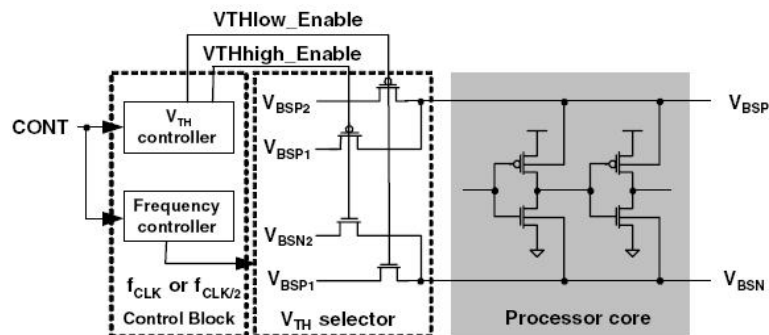
Слика 2.7: DVTS шема за одређивање оптималне вредности напона прага транзистора

PMOS и NMOS транзистора у одговарајућу фреквенцију осциловања. Сигнал грешке који представља разлику између референтне фреквенције и фреквенције

са излаза VCO осцилатора води се на улаз контролера повратне спреге. Помоћу овог контролера одређује се фреквенција пумпи наелектрисања.

DVTS техника која је ефикасана у минимизацији статичке потрошње углавном се користи у комбинацији са DVS техником која ефикасно минимизира динамичку потрошњу [KSM+98, Kur04, MA11]. У том случају уштеде енергије могу бити и до 65% [MA11].

Простији метод за контролу напона прага код кога се V_{th} мења динамички и то између две могуће вредности (високе и ниске), у зависности од тренутних захтева за брзином рада система приказан је на слици 2.8 [NHK+01]. За разлику од претходно описаног метода код кога напон прага транзистора може да се мења континуално, овде постоје две могуће дискретне вредности за V_{th} . Уз помоћ *CONT* сигнала који се добија из процесора (тј. софтверски се контролише [LS00]) блок V_{th} контролера сетује *VTHlow_Enable* или *VTHhigh_Enable* сигнал који даље управља поларисањем субстрата процесора тј. одређује напон прага транзистора. Са сетованим *VTHlow_Enable* сигналом транзистори у систему имају директно поларисани субстрат тј. ниску вредност напона прага. У случају да је могуће правити компромисе између брзине рада и потрошње сетује се *VTHhigh_Enable* сигнал. Транзистори у систему сада имају већу вредност напона прага што утиче на смањивање брзине рада али и статичке потрошње. *CONT* сигнал такође контролише и *Frequency controller* блок тј. радну фреквенцију система. Кад транзистори у систему имају мању вредност напона прага радна фреквенција система је f_{CLK} . У случају већих вредности напона прага транзистора радна фреквенција је дупло мања. Аутори су у [NHK+01] ову технику адаптивних промена напона прага V_{th} у зависности од тренутно потребне брзине рада применили на MPEG-4 систем видео кодирања. Уштеде у потрошњи енергије могу бити и до 27%. На истом примеру показали су да се увођењем и треће дискретне вредности за напон прага V_{th} систем у погледу потрошње незнатно побољшава (само 6%).



Слика 2.8: Шема динамичке контроле напона прага транзистора

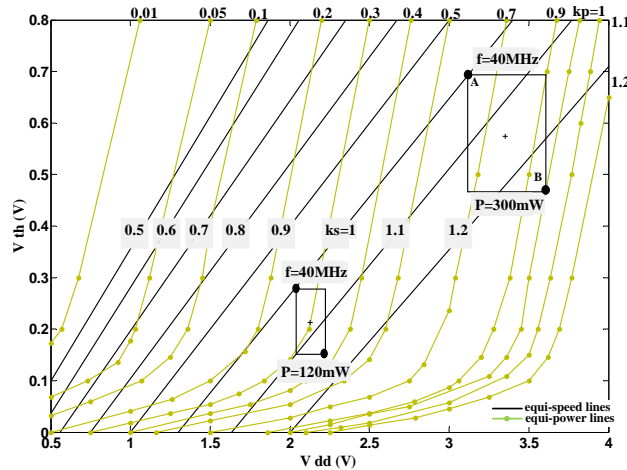
2.1.4. Мулти V_{dd}/V_{th} технике за минимизацију потрошње

Као што је већ показано, напон напајања система и напон прага транзистора имају знатан утицај како на перформансе тако и на потрошњу електричне енергије. На слици 2.9 у $V_{dd}-V_{th}$ равни приказане су линије константе потрошње (equi-power) и брзине (equi-speed) добијене на основу алфа модела транзистора [SN90]. Претпоставимо, на пример, да нам је технолошки процес наметнуо следећа

ограничења: $V_{dd}=3.3V \pm 10\%$ и $V_{th}=0.55V \pm 0.1V$. Простор дозвољених вредности за ове две величине представљен је већим правоугаоником на слици 2.9.

Сваки дизајн чија су ова два параметра унутар овог правоугаоника требало би да испуњава све задате спецификације у погледу својих перформанси. Ако се параметри V_{dd} , V_{th} дизајна налазе у углу А већег правоугаоника дизајн ће имати највеће кашњење тј. биће најспорији. У углу В потрошња електричне енергије дизајна биће максимална. Бољи компромиси између потрошње и брзине рада дизајна могу се постићи смањивањем флукуација тј. толеранција напона напајања и напона прага, и то нарочито у области мањих вредности за V_{dd} . Линије константне брзине и потрошње нормализоване су у тачкама А и В нормализационим факторима k_s и k_p , респективно. То нам омогућава да лако видимо како се мењање положаја и величине правоугаоника у V_{dd} - V_{th} равни одражава на брзину рада и потрошњу дизајна. На пример, за $V_{dd}=2.1V \pm 5\%$ и $V_{th}=0.18V \pm 0.05V$ (мањи правоугаоник на слици 2.9) потрошња електричне енергије мања је за 60% ($k_p=0.4$), док брзина рада остаје непромењена ($k_s=1$).

У [PSL06] аутори закључују да од свих могућих комбинација V_{dd}/V_{th} са којима ће систем да испуни задата временска ограничења само једна комбинација гарантује минималну потрошњу електричне енергије. Положај те оптималне радне тачке $\{V_{ddopt}, V_{thopt}\}$ као и њена потрошња чврсто су везани са технолошким параметрима и архитектуром дизајна.



Слика 2.9: Линије константне потрошње и брзине у V_{dd} - V_{th} равни

Иста група аутора у [SNP+06] даје једначину (базирану на алфа моделу транзистора) за прорачун укупне потрошње дизајна који ради на оптималној радној тачки $\{V_{ddopt}, V_{thopt}\}$. Укупна потрошња представљена је као збир динамичке и статичке потрошње која је последица само струје цурења у каналу:

$$P_{tot}^{opt} \cong \frac{aCNf}{(1-\chi A)^2} \left[SV_T \left(\ln \left(\frac{I_0}{2SV_T a C f} (1-\chi A) + 1 \right) + \chi B \right) \right]^2 \quad (2.1)$$

где је N број ћелија у колу, a просечна активност ћелија (тј. број ћелија које мењају своје стање у односу на укупан број ћелија), C еквивалентна капацитивност ћелије, V_T термални напон, f радна фреквенција, I_0 просечна off-струја по ћелији за $V_{GS}=V_{th}$ а S нагиб у режиму слабе инверзије. A и B су две променљиве које служе за

фитовање и зависе од константе α из алфа модела транзистора. Променљива χ једнака је:

$$\chi^\alpha = \frac{\xi \cdot f \cdot LD}{I_0 \left(\frac{e}{\alpha S V_T} \right)^\alpha} \quad (2.2)$$

са променљивом ζ (мереном у фарадима) која такође представља параметар за фитовање и у себи укључује капацитивност гејта који свичује (мења стање) и параметром LD (Logical Depth) који представља кашњење критичне путање.

Једначина 2.1 је врло битна због тога што нам омогућава да аналитички проценимо оптималну укупну потрошњу директно преко параметара који зависе од архитектуре дизајна (активност (a), број ћелија (M), фреквенција (f), логичка дубина (LD)) као и од технолошких параметара (просечна off-струја I_0 , нагиб у режиму слабе инверзије (S), коефицијент α из алфа модела транзистора итд.). Такође, почевши од једначине 2.2 могуће је разумети утицај уобичајених трансформација архитектуре дизајна (паралелизам, проточност) као и поредити перформансе једне исте архитектуре са различитим технолошким параметрима.

У [NS00b] аутори представљају формуле за одређивање оптималних (са становишта минималне потрошње) вредности напона напајања и напона прага за задате технолошке параметре дизајна као и за задата временска ограничења. Ове формуле базирају се на алфа моделу транзистора а у обзир узимају и ефекте кратког канала транзистора, варијације напона прага као и температуру.

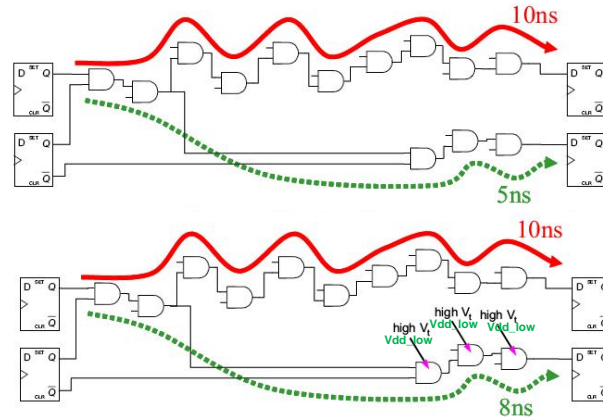
И овде треба нагласити да дизајнери интегрисаних кола углавном немају могућност да континуално мењају напоне напајања и напоне прагова транзистора. Најчешћи случај је да дизајнери могу да бирају између неколико дискретних вредности за V_{dd}/V_{th} тј. да у библиотеци стандардних ћелија имају на располагању неколико типова ћелија (Ultra High Speed – UHS, High Speed – HS, Low Leakage – LL итд.) које имају предефинисане V_{dd} односно V_{th} . У [НОКО1] дат је оптималан (са становишта потрошње електричне енергије) међусобни однос V_{dd} (односно V_{th}) у случају да су на располагању од две до четири дискретне вредности за ове величине (слика 2.10).

$3a \{V_{dd1}, V_{dd2}\}:$	$\frac{V_{dd2}}{V_{dd1}} = 0.5 + 0.5 \frac{V_{th}}{V_{dd1}}$	$3a \{V_{th1}, V_{th2}\}:$	$V_{th2} = 0.10V_{dd} + V_{th1}$
$3a \{V_{dd1}, V_{dd2}, V_{dd3}\}:$	$\frac{V_{dd2}}{V_{dd1}} = \frac{V_{dd3}}{V_{dd2}} = 0.6 + 0.4 \frac{V_{th}}{V_{dd1}}$	$3a \{V_{th1}, V_{th2}, V_{th3}\}:$	$V_{th2} = 0.06V_{dd} + V_{th1}$ $V_{th3} = 0.07V_{dd} + V_{th2}$
$3a \{V_{dd1}, V_{dd2}, V_{dd3}, V_{dd4}\}:$	$\frac{V_{dd2}}{V_{dd1}} = \frac{V_{dd3}}{V_{dd2}} = \frac{V_{dd4}}{V_{dd3}} = 0.7 + 0.3 \frac{V_{th}}{V_{dd1}}$	$3a \{V_{th1}, V_{th2}, V_{th3}, V_{th4}\}:$	$V_{th2} = 0.04V_{dd} + V_{th1}$ $V_{th3} = 0.05V_{dd} + V_{th2}$ $V_{th4} = 0.06V_{dd} + V_{th3}$

Слика 2.10: Однос напона напајања и напона прага у мулти V_{dd}/V_{th} дизајну

Потрошња електричне енергије применом мулти V_{dd}/V_{th} техника минимизира се тако што се делује на транзисторе у некритичним путањама кола. Повећавање напона прага или смањивање напона напајања (или пак примена обе технике у исто време) у некритичним путањама повећаће кашњење на њима али неће деградирати перформансе система (види слику 2.11). Са друге стране смањиће се потрошња, како статичка (услед повећања V_{th}) тако и динамичка (услед смањивања V_{dd}). Бројне студије су показале да значајне уштеде (почев од мање од 10% па до скоро 50%, у просеку 40%) у динамичкој потрошњи могу да проистекну

применом мулти V_{dd} технике [UI00, DMM+02]. Колико ће се енергије уштедети грубо зависи од два параметра: разлике између регуларног напона напајања V_{dd} и мањих вредности за напон напајања $V_{dd,low}$ као и од процената транзистора који се напајају мањим вредностима V_{dd} . Треба рећи и то да бенефит од употребе више напона напајања брзо улази у засићење и да се највеће уштеде потрошње дешавају приликом преласка са једног на два различита напона напајања [НОК01]. Увођење трећег и сваког даљег извора за напајања само незнатно оптимизује потрошњу. Такође, технолошки напредак мулти V_{dd} технику чини мање ефикасном (са преласком од 250nm до 90nm ефикасност ове методе се дупло смањује [Mac04]).



Слика 2.11: Мулти V_{dd}/V_{th} техника примењена на транзисторе у некритичној путањи

Да би дизајн на који се примењује мулти V_{dd} техника исправно функционисао неопходно је и присуство конвертора нивоа – компоненте која ће бити нека врста интерфејса између делова у дизајну који се напајају различитим вредностима V_{dd} . Присуство конвертора нивоа повећава површину и кашњење дизајна а такође и смањује ефикасност минимизационе технике будући да и сам троши одређену енергију за конверзију. Неке верзије овог конертора могу се пронаћи у [ISN04, MR04, PCK+08].

Мулти V_{th} техника за минимизацију статичке потрошње исту може да редукује од 50 па чак и до 80%. Недостатак мулти V_{th} технике је тај што захтева додатне маске (за добијање транзистора са различитим напонима прага) што утиче на повећање трошкова производње. Такође, треба применити и специјална правила дизајнирања на границама између области у којима су транзистори са различитим напонима прагова [Mac04].

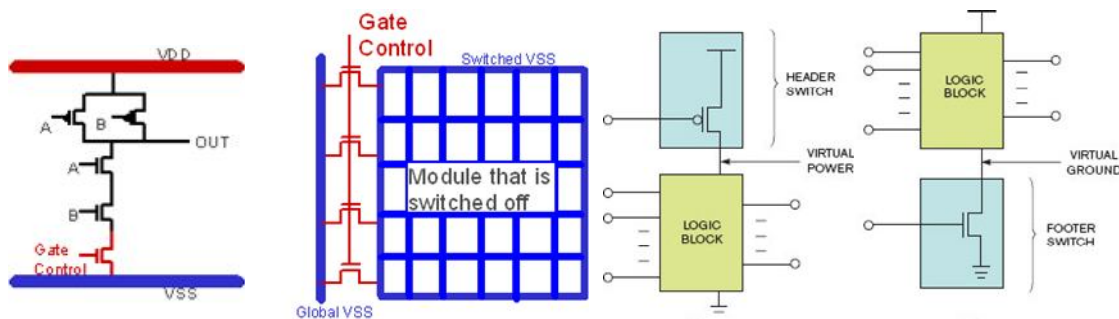
2.1.5. Гејтовање напона напајања и такта

Велике струје цурења у *standby* режиму рада (а самим тим и статичка потрошња) један су од главних недостатака многих батеријски напајаних мобилних уређаја. Узрок ових струја су ниски напони прагова транзистора. Са друге стране, смањивање напона прага транзистора је нужност и врши се услед константног смањивања напона напајања, како би се одржале жељене брзине рада тј. процесирања података. Један од начина смањивања цурећих струја неактивних компоненти у дизајну је техника гејтовања напона напајања. Уз помоћ специјалних хедер (PMOS) односно футер (NMOS) транзистора-прекидача логички

блокови у дизајну који су у датом тренутку неактивни једноставно се дисконектују са мреже за напајање [CCC+06, CLT+10]. Због мање отпорности у проводном стању хедер PMOS транзистори чешће су у употреби [KR09]. Неке од варијанти гејтовања напона напајања приказане су на слици 2.12.

Код финог гејтовања транзистор-прекидач енкапсулира се у саму стандардну ћелију. На слици 2.12а) приказана је двоулазна NAND ћелија са *Gate Control* футер транзистором. Величина овог транзистора прорачунава се према најгорем случају у коме се претпоставља да ће ћелија (логичко коло) да мења своје стање у сваком тактном циклусу. Због тога овај транзистор-прекидач заузима доста површине што представља главни недостатак ове врсте гејтовања. Како би и сам имао мале цуреће струје транзистор прекидач обично је са великим напоном прага. Техником финог гејтовања статичка потрошња може се редуковати и до 10 пута, а могуће је и примењивати је заједно са другим техникама за минимизацију [LCG+09].

Код грубог гејтовања транзистори-прекидачи су део дистрибуционе мреже напајања (слика 2.12б). Две најчешће структуре грубог гејтовања су прстенаста и решеткаста [JD10]. Будући да је у сваком тренутку активан само један део кола тј. само неки транзистори-прекидачи њихова величина углавном је мања у поређењу са транзисторима-прекидачима код финог гејтовања.



(a)

(б)

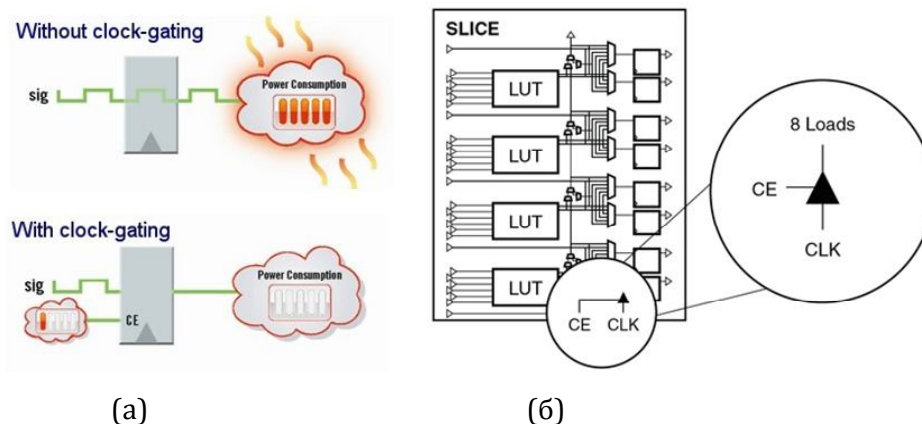
Слика 2.12: Гејтовање напона напајања: (а) фино гејтовање, (б) грубо гејтовање

Један од изазова који се јавља приликом гејтовања напона напајања је решавање проблема памћења стања секвенцијалних кола. Ако се у тренутку искључења са мреже за напајање секвенцијалним колима не запамти стање систем неће исправно функционисати. Проблем памћења стања ових кола може се решити помоћу постојаних (non-volatile) меморија или специјалних флип-флопова за памћење стања [MR04, HEG+06]. Пре него што се искључе са мреже за напајање, стања секвенцијалних кола упишу се у меморију или флип-флопове одакле се ишчитавају након поновног повезивања на мрежу. Због много веће брзине рада употреба флип-флопова је далеко чешћа.

Будући да транзистори-прекидачи показују одређену инертност, тј. да постоји одређени степен латенције приликом сваког „успављивања“ и „буђења“ дела кола које се гејтује као и додатне енергије која се утроши приликом сваког „буђења“, аутори су у [ADS+06] представили технику гејтовања са више нивоа *sleep* режима. Сваки *sleep* режим ниво представља компромис између латенције/енергије буђења и уштеде статичке потрошње. Што је већа латенција буђења потребно је више енергије да се део кола који је претходно дисконектован са мреже на исту

поново конектује. Са друге пак стране, уштеде статичке потрошње су веће. У односу на конвенционално гејтовање напона напајања, применом технике више нивоа *sleep* режима аутори су у [ADS+06] уштедели додатних 17% статичке потрошње.

На сличан начин, гејтовањем такта код синхроних кола, могуће је редуковати динамичку потрошњу неког дизајна. Помоћу сигнала дозволе такта (*clock enable - CE*) блокира се тактовање делова кола који су у одређеном тренутку неактивни и тиме спречава њихова нежељена транзициона активност тј. минимизира се динамичка потрошња (слика 2.13а). Потенцијали ове технике за минимизацију динамичке потрошње препознати су и од стране Xilinx-а, једног од највећих произвођача комерцијалних програмабилних FPGA компоненти. Xilinx је у својој ISEv12 верзији CAD софтвера додао дизајнерима могућност оптимизације помоћу интелигентног гејтовања такта. Колико ће се динамичке потрошње енергије применом ове методе уштедети зависи пре свега од типа дизајна који се имплементира. Инжењери из Xilinx-а су гејтовање такта применили на сет дизајна велике практичне примене имплементираних на FPGA компоненту из Virtex-6 фамилије. Просечна забележена уштеда динамичке потрошње била је око 30%. Највећа уштеда (41%) забележена је на дизајну који је обављао 3DES енкрипцију података. Логика за гејтовање такта незнатно (2%) је повећавала површину дизајна [Riv10]. На слици 2.13б) приказана је унутрашња структура Xilinx Virtex-6 slice-а.



Слика 2.13: (а) Техника гејтовања такта, (б) структура Xilinx Virtex-6 FPGA slice-а

2.1.6. „Гомилање“ транзистора

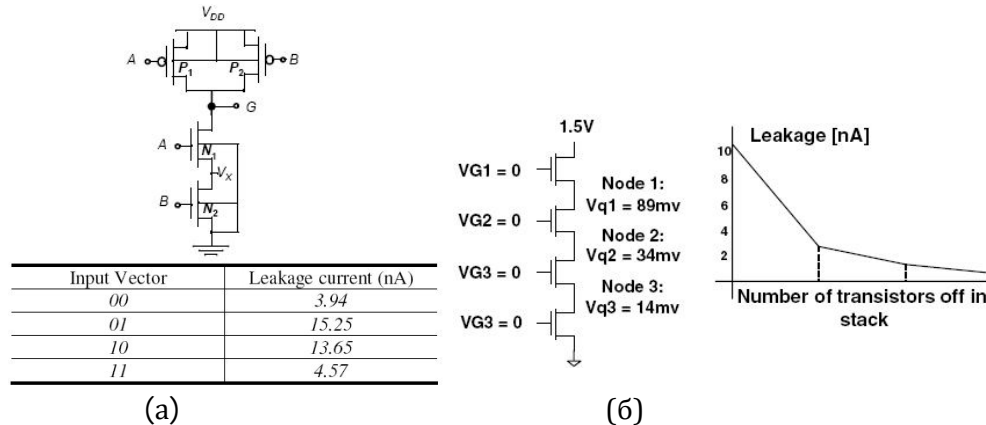
„Гомилање“ транзистора представља ефикасну технику за минимизацију струје цурења у каналу која се експоненцијално повећава са смањивањем напона прага транзистора и постаје једна од главних претњи смањивању аутономије батерија код батеријски напајаних уређаја. У [HN97, YBD98, NBD+01] показано је да „гомилање“ тј. слагање два или више транзистора који су у *off* стању знатно редукује струју цурења у каналу у односу на један *off* транзистор. Објашњење ове појаве састоји се у следећем: на слици 2.14а) приказано је једно NAND коло код кога постоје два нагомилана NMOS транзистора, N_1 и N_2 . У случају да су оба транзистора у *off* стању ($AB=“00“$) напон у тачки између два транзистора (V_X) је

позитиван услед мале струје дрејна. Позитиван потенцијал у овој тачки има следеће ефекте:

- Напон између гејта и сорса транзистора N_1 (V_{gs1}) постаје негативан. Као последица тога струја цурења у каналу експоненцијално се смањује.
- Потенцијал између субстрата и сорса транзистора N_1 (V_{bs1}) постаје негативан. Као последица тога повећава се његов напон прага V_{th} што такође утиче на експоненцијално смањивање струје цурења у каналу.
- Напон између дрејна и сорса транзистора N_1 (V_{ds1}) се смањује. Као последица овога повећава се напон прага транзистора (услед смањивања DIBL ефекта) што поново смањује струју цурења у каналу.

Што је више нагомиланих транзистора у *off* стању то су већи напони у тачкама између транзистора (напони на сорсовима транзистора) а самим тим и већа редуција струје цурења у каналу тј. статичке потрошње. На слици 2.14б) приказан је утицај броја нагомиланих транзистора на струју цурења у каналу [Fig05]. Највећа редуција цуреће струје (скоро за ред величине) дешава се кад се уместо једног транзистора нагомилују два. Може се, дакле, закључити да редуција статичке потрошње зависи од улазног вектора тј. комбинације која се доводи на гејтове транзистора [CWJ+98]. На слици 2.14а) приказана је зависност струје цурења код 2-улазног NAND кола пројектованог у 130nm CMOS технологији од улазног вектора.

Појава све већих струја цурења оксида гејта као последице сталног смањивања димензија транзистора негативно утиче на ефекте „гомилања“ транзистора [MNC+03].



Слика 2.14: (а) Ефекти гомилања транзистора у NAND колу, (б) утицај броја нагомиланих транзистора на струју цурења у каналу

2.1.7. Оптичке интерконеције на чипу

Све до 1997. године електричне везе за повезивање транзистора унутар интегрисаног кола као и мреже напајања и такта биле су алуминијумске. Након тога је, најпре у IBM-у, алуминијум био замењен бакром који се истицао скоро дупло мањом отпорношћу, већом тачком топљења, већом поузданошћу и бољом отпорношћу на електромиграције [Ibm97]. Од тада је сталан тренд смањивања димензија транзистора и повећања њиховог броја унутар једног интегрисаног

кола утицао на то да ове везе за интерконекције постану све краће и тање (што је повећавало њихову отпорност), и да при том преносе све веће количине наелектрисања и то на већим удаљеностима, како би се снабдели сви транзистори. Због тога су почели да се јављају неки нежељени ефекти као што су повећана динамичка потрошња, латенција тј. кашњење сигнала који се простире кроз везу за повезивање, преслушавање између интерконекција, пад напона на вези, повећан шум итд. На повећање латенције утицао је и слој диелектрика који се умеће између два слоја за повезивање. Чак и употреба материјала мале диелектричне константе (low-k) није довољна да се временска RC константа интерконекције смањи на жељени ниво. Због свега поменутог, један од ограничавајућих фактора за постизање жељених брзина рада код савремених интегрисаних кола постаје и интерконекција тј. веза између гејтова. Стога је неопходна даља иновација материјала који се користе за повезивање или примена неких нових, неконвенционалних техника повезивања транзистора унутар интегрисаног кола.

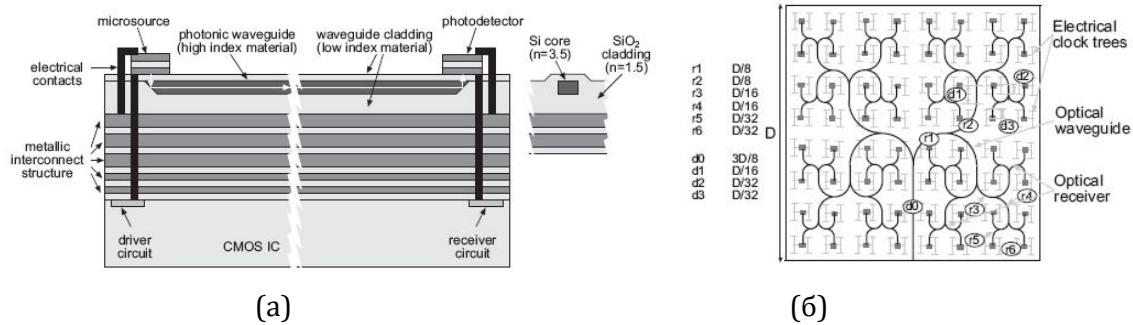
У [Ali06] аутор представља студију у којој се предлаже да се бакар замени карбонским наноцевима. Главни аргументи за прелазак на ове цеви су балистички транспорт наелектрисања (тј. много већа покретљивост носилаца наелектрисања у односу на бакар), могућност преноса знатно веће количине наелектрисања и већа термална проводност.

Бакарне везе могуће је заменити и увођењем слоја оптичких интерконекција на чипу. У [Mil96, VTJ+99] показано је да оптичке интерконекције конзумирају мање енергије у односу на електричне. Поред мање потрошње, велике предности оптичке интерконекције у односу на електричну су и повећана процесна моћ, повећана имуност на електромагнетне сметње и преслушавање између канала, велика синхронизованост унутар самог кола као и са другим колима, смањена осетљивост на температурне варијације итд. Велики изазов за увођење слоја оптичких интерконекција на чипу је чињеница да процес његове фабрикации треба да буде компатибилан са досадашњим и будућим CMOS технолошким процесима. Такође, потребно је решити и проблем ефикасности конверзије електричног сигнала у оптички и обрнуто [HCC+06].

Хибридна (3D) интеграција оптичког слоја на врх CMOS интегрисаног кола приказана је на слици 2.15а). Оваква структура интегрисаног кола реално се може очекивати у будућности. Микроласер (описан у [FSB99]) је повезан са структуром пасивног таласовода и обезбеђује сигнал оптичком пријемнику. На страни пријемника, оптички сигнал велике брзине конвертује се у електрични сигнал и даље дистрибуира локалном електричном мрежом интерконекција. Планарни оптички таласовод изграђен је од силицијума, Si са облогама од силицијум диоксида SiO₂ (слика 2.15а) [PGH+04].

Електричну дистрибуциону мрежу такта такође је могуће заменити оптичком. У [PGH+04] вршено је поређење потрошње електричне енергије два истоветна дизајна пројектована у 70nm CMOS технологији. Једина разлика међу дизајнима била је мрежа дистрибуције такта: један дизајн био је са електричном а други са оптичком дистрибуционом мрежом такта приказаном на слици 2.15б). Док је потрошња дизајна са електричном дистрибуционом мрежом такта зависила од радне фреквенције, потрошња дизајна са оптичком мрежом такта била је фреквентно независна и око 5 пута мања на фреквенцији од 5GHz.

Једна од техника која ће се у будућности сигурно користити је и техника такозваног мултиплексирања таласне дужине (WDM – wavelength division multiplex). Примена ове технике омогућиће да се, уместо преко магистрале података, кроз један исти оптички слој, у истом тренутку преноси више сигнала, и то сваки од њих на различитој таласној дужини. Примена ове технике такође ће омогућити и истовремену бидирекциону комуникацију [AS02].



Слика 2.15: (а) Хибридна интеграција оптичког слоја, (б) Оптичка мрежа Н-стабла такта са 64 излазна чвора

2.1.8. Хибридни приступи у минимизацији потрошње

Свакако да ће правилна примена две или више различитих техника за минимизацију потрошње електричне енергије дати боље резултате од примене само једне минимизационе технике. Треба такође рећи и то да не постоји неко „златно правило“ о броју, врсти и редоследу примењивања минимизационих техника. Колико је енергије могуће уштедети (и да ли је могуће жртвовати перформансе у замену за мању потрошњу) зависи искључиво од технолошких карактеристика интегрисаног кола као и од типа функције коју то коло обавља. Много је радова у литератури који се баве анализом међусобног утицаја претходно описаних минимизационих техника на потрошњу електричне енергије.

О међусобном утицају напона напајања V_{dd} и напона прага V_{th} на минимизацију потрошње могуће је прочитати у [NS00b, SS03].

Међусобни утицај напона прага V_{th} и димензионисања транзистора (мењање ширине канала W или дебљине оксида гејта t_{ox}) на статичку или укупну потрошњу истраживан је у [KYT+02, NDO+03, LDB+04]. Резултати показују да је код неких дизајна могуће уштедети и до 57% електричне енергије [NDO+03].

На крају, о међусобном утицају три или више минимизационих техника могуће је информисати се у [НОК01, Aug02, ВНМ+02, SSB04, НХV+04]. У [Aug02] дата је исцрпна анализа међусобног утицаја мулти (dual) V_{dd} , мулти (dual) V_{th} , и минимизационе технике која мења ширину канала транзистора. Испитиван је утицај сваке минимизационе технике на типичне логичке блокове, засебно и у комбинацији са осталим техникама. Такође је вођено рачуна и о редоследу примењивања ових техника (коришћено је 15 различитих комбинација). Закључак је да не постоји апсолутно најбоља минимизациона техника нити апсолутно најбољи редослед примењивања техника. У случају да дизајн на који се

примењују технике има велику транзициону активност (тј. динамичку потрошњу), најбоље је прво применити дуал V_{dd} технику па након тога димензионисање транзистора и на крају дуал V_{th} . Са друге стране, уколико у дизајну доминира статичка потрошња (ниска транзициона активност) дуал V_{th} техника има предност у односу на дуал V_{dd} и димензионисање транзистора. У [ВНМ+02] аутори закључују да потенцијалне уштеде енергије у неком дизајну зависе од његовог енергетског профила. Такође, максимална уштеда енергије дешава се онда када су потенцијали свих променљивих које могу да утичу на потрошњу (напон напајања, напон прага, димензије транзистора) међусобно избалансирани. За сваку од ових променљивих аутори су прорачунали осетљивост енергије у односу на кашњење и тиме повезали њихове потенцијале уштеде енергије са физичким својствима дизајна. Применом ове оптимизационе технике на ланац инвертора, SRAM декодере и сабираче аутори су проценили да је, уз жртвовање 20% перформанси система, могуће уштедети 40 до 70% енергије. Без жртвовања перформанси (повећања кашњења) могуће су уштеде енергије и до 50%.

2.1.9. CMOS алтернативе за минимизацију потрошње

У [SPV+11] аутори су представили студију о томе како техника напрезања силицијума може да се искористи за минимизацију потрошње. Ову технику комбиновали су са dual V_{th} техником и димензионисањем транзистора. Уштеда енергије остварује се тако што се у гејтовима у дизајну који имају велику транзициону активност врши напрезање силицијума (ради повећања покретљивости носилаца наелектрисања) чиме се омогућава додатно смањивање димензија транзистора и уштеда динамичке потрошње. Наравно, као последица смањивања димензија транзистора повећава се статичка потрошња али у мањој мери него што се штеди динамичка. Са друге стране, транзистори у гејтовима који имају ниску транзициону активност су са великим напонима прагова. На тај начин им се редукује статичка потрошња а компензација смањивања перформанси врши се повећањем њихових димензија. Повећање димензија транзистора повећава њихову динамичку потрошњу али је и овде, на сличан начин као код гејтова са великом транзиционом активношћу, уштеда у статичкој потрошњи већа у односу на губитак у динамичкој што води ка минимизацији тоталне потрошње. Уз побољшање перформанси дизајна од 15% аутори пријављују и 5.4% уштеде енергије.

У [PL11] представљена је програмабилна и скалабилна техника за дизајнирање кола на бази спинтроничких (MTJ) компоненти које су у стању да имплементирају било коју двоулазну логичку функцију ниске потрошње електричне енергије.

Уз констатацију да је топлота која се ослобађа приликом функционисања интегрисаног кола један од најгорих непријатеља електронских компоненти данашњице и да ће наредних година око 80% укупне потрошње микропроцесора бити конзумирано од стране интерконекција, у областима на чипу које је тешко хладити, у [Bal09] описана су алтернатива конвенционалном повезивању транзистора – употреба материјала са високом термалном проводношћу (графена, дијамант, угљеничне наноцеви итд.), као и неке врсте неконвенционалног хлађења (термоелектрично и водено хлађење).

У [PGH+04] показано је како неке нано компоненте (транзистор са једним електроном (SET), карбонске наноцеви), засебно или у комбинацији са конвенционалним CMOS структурама, могу да редукују потрошњу електричне енергије. Главна предност оваквих компоненти су њихове мале димензије и изузетно мала количина наелектрисања (чак до једног електрона) потребна да се кодира стање логичке јединице.

На крају, треба поменути и *pass-transistor* логику, као популарну и широко прихваћену алтернативу CMOS транзисторима [RCN03]. Главна идеја код *pass-transistor* логики је да се редукује број транзистора потребних да се имплементира нека логичка функција и то тако што ће бити дозвољено да се примарни улази воде не само на гејтове већ и на сорсеве и дрејнове транзистора [RWM85]. Редукција броја транзистора даље води ка смањеној капацитивности а самим тим и смањеној потрошњи електричне енергије. Уштеда у потрошњи електричне енергије долази и из чињенице да *pass* транзистори захтевају мање енергије за пуњење капацитивности чворова. Деградирање нивоа логичке јединице ($V_{dd}-V_{th}$) на излазу *pass* транзистора један је од главних недостатака овакве транзисторске структуре зато што спречава каскадно везивање гејтова на бази *pass* транзистора (повезивање излаза једног *pass* гејта на гејт улаз другог *pass* транзистора). Једно од решења проблема деградирања нивоа логичке јединице је његова рестаурација уз помоћ PMOS транзистора у петљи повратне спреге [RCN03]. Проблем се такође може решити увођењем транзистора са нултим напоном прага ($V_{th}=0V$). Ипак, једно од најчешћ коришћених решења елиминације утицаја деградирања нивоа логичке јединице је логика трансмисионих гејтова – структура које се састоје од паралелно везаних NMOS и PMOS транзистора. Неки од покушаја минимизације потрошње уз помоћ *pass* транзисторске логики могу се наћи у [OIT02, SS02].

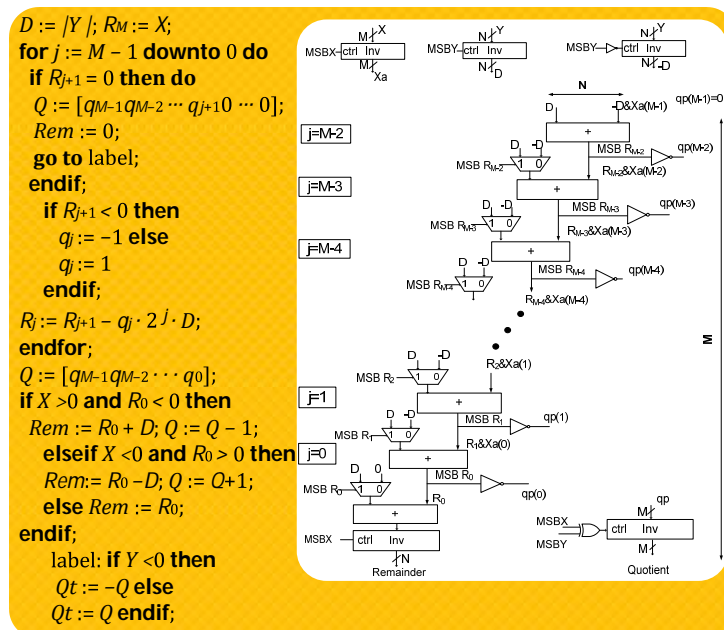
2.1.10. Демонстрација утицаја технолошких параметара на потрошњу

Ради практичне демонстрације утицаја технолошких параметара на потрошњу електричне енергије референтни дизајн 12-битног бинарног делитеља описан је у VHDL програмском језику и имплементиран на Xilinx-ове FPGA компоненте различитих фамилија (Virtex-4, Virtex-5, Virtex-6, Virtex-6 Lower Power). За бинарно дељење коришћен је *Radix-2 non-restoring* алгоритам [JC11] са целобројним означеним количником и остатком. Програмска секвенца дељења заједно са хардверском структуром делитеља приказана је на слици 2.16.

Имајући, у општем случају, на улазу дељеник X и делилац Y као M - и N -то битне целе означене бројеве представљене у другом комплементу, респективно, алгоритам дељења на излазу делитеља даје количник Qt и остатак Rem . Количник и остатак су такође M - и N -то битни цели означени бројеви представљени у другом комплементу, тако да је испуњено $X=Y*Qt+Rem$. Остатак Rem је истог знака као и дељеник X . Више детаља о *Radix-2 non-restoring* бинарном делитељу читалац може пронаћи у поглављу 4.5.1 ове дисертације.

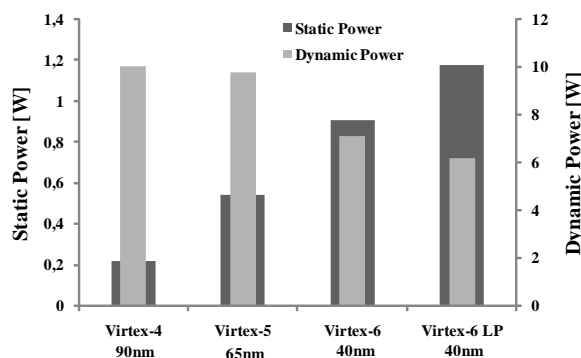
За мерење потрошње имплементираних дизајна коришћен је XPower Analyzer алат у оквиру ISE 12.4 програмског пакета. Улази делитеља генерисани су у MATLAB-у, као сигнали са Gauss-овом расподелом (по 1000 вредности за дељеник

и делилац). Средња вредност, аутокорелација и међусобна корелација ових сигнала била је једнака нули. Резултати имплементације приказани су на слици 2.17.



Слика 2.16: Radix-2 non-restoring алгоритам бинарног дељења

Различити технолошки процеси (од 90nm до 40nm), варијације напона напајања (од 1.2V до 0.9V), напона прага и димензија транзистора очигледно утичу на потрошњу електричне енергије, како статичку тако и динамичку. Очигледан је и тренд пораста статичке потрошње са преласком на сваку нову генерацију FPGA компоненти. Пораст статичке потрошње углавном се дешава услед пораста цурећих струја које су последица смањивања транзисторских димензија. Краћи транзисторски канали и тањи оксиди гејта (који се углавном користе при преласку на нову генерацију) омогућавају лакше протицање цурећих струја, било у региону канала транзистора, било кроз оксид гејта.

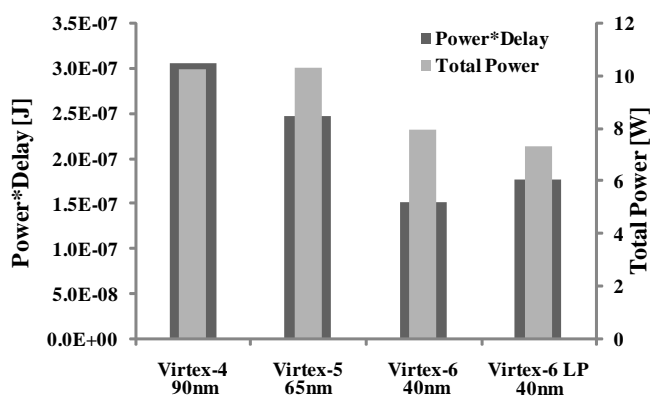


Слика 2.17: Статичка и динамичка потрошња референтног дизајна бинарног делитеља имплементираних на FPGA компоненте различитих технолошких карактеристика

Што се тиче динамичке потрошње, напон напајања језгра FPGA компоненте као и капацитивност чвора генерално се смањују са сваком новом генерацијом

обезбеђујући тиме значајне уштеде у динамичкој потрошњи у односу на претходну генерацију. Инжењери из Xilinx-а тврде да ће процес производње силицијума прилагођен FPGA компонентама и иновативна јединствена архитектура омогућити њиховим новим 28nm FPGA компонентама из Virtex-7 фамилије 50% уштеде укупне потрошње у односу на претходну 40nm генерацију [Xil11].

Последњих година производ потрошње и кашњења (power-delay product – PDP) постаје мерило ефикасности или ефективности дигиталног интегрисаног кола у погледу кашњења и потрошње. Мањи PDP значи да се потрошња боље „преводи“ у брзину рада дигиталног интегрисаног кола. Такође, мањи PDP може да значи и да ће повећање кашњења кола да има већи ефекат у погледу смањивања његове потрошње, будући да су брзина рада и потрошња дизајна међусобно опречни захтеви. Са слике 2.18 види се да Virtex-6 фамилија FPGA компоненти (не Virtex-6 Lower Power – LP као што би неко могао да очекује) има најбољи (најмањи) PDP. Објашњење лежи у чињеници да је, у поређењу са Virtex-6 фамилијом, једина измена у Virtex-6 LP смањени напон напајања језгра (од 1V на 0.9V). Смањивање напона напајања језгра очигледно смањује динамичку потрошњу (12.5%, види слику 2.17). Са друге стране, побољшање у погледу потрошње „преплаћено“ је смањивањем брзине рада од 26.6% као што се види из табеле 2.1.



Слика 2.18: PDP производ и укупна потрошња сета FPGA фамилија

Табела 2.1: Пропагационо кашњење бинарних делитеља

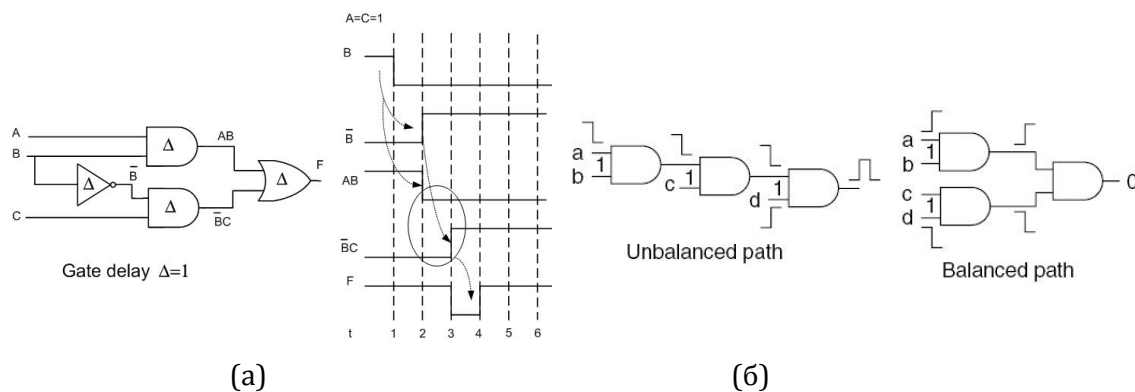
	Virtex-4	Virtex-5	Virtex-6	Virtex-6 LP
Delay [ns]	29.83	23.98	19.04	24.11

2.2. Оптимизација на блоковском нивоу

Обавезан део тока дизајнирања сваког интегрисаног кола (ASIC, FPGA, custom IC) представља процес физичке имплементације. Овај процес је обично итеративан и подељен у неколико фаза. У свакој итерацији, у фази размештања (placement) дизајн се најпре подели у групе или блокове, довољно мале да могу да се сместе у једну „градивну јединицу“ (ASIC стандардну ћелију, FPGA конфигурабилни логички блок – CLB (Configurable Logic Block) итд.). Након тога, овим јединицама се додељује одређена физичка локација на чипу. На крају, размештени блокови се у фази рутирања (routing) међусобно повезују електричним везама. Одређивање

путање веза приликом рутирања се обично обавља у две фазе. Након грубог или глобалног размештаја, у фази детаљног повезивања свакој вези додељује се јединствена комплетна путања. У случају ASIC чипа или неког интегрисаног кола по поручбини (custom IC), након фазе размештања и повезивања могуће је генерисати маске за фабрикацију и почети процес производње интегрисаног кола.

У свакој фази процеса физичке имплементације дизајна тежи се оптимизацији његових коначних перформанси (минимизирању површине чипа, потрошње и кашњења), али тако да се не угрожава изводљивост следећих фаза у процесу. Главни акценат у процесу размештања и повезивања (Placement And Routing – PAR) је на минимизацији дужина веза за међусобно повезивање будући да се дужина веза директно пресликава на време потребно да се сигнал пренесе са једног места на друго тј. на брзину рада целокупног система. Такође, минимална дужина веза за међусобно повезивање води ка мањој капацитивности интерконекија а самим тим и ка мањој динамичкој потрошњи дизајна. У току размештања и повезивања тежи се, такође, и елиминацији гличева, нежељених активности (појава „лажних стања“) у чворовима интегрисаног кола које повећавају транзициону активност а самим тим и динамичку потрошњу. На слици 2.19а) приказан је један пример појаве глича на излазу логичког кола. Ако су сигнали A и C оба у стању логичке '1' излаз кола би такође требало да је у стању логичке '1', без обзира на стање сигнала B . Ипак, услед кашњења на инвертору, може се видети да су у једном кратком временском интервалу (једнаком кашњењу сигнала кроз инвертор) оба излаза логичких AND кола у стању логичке '0' што узрокује глич на излазу (појаву „лажне“ '0'). Главни узрок појаве гличева су неизбалансиране путање сигнала тј. њихова различита времена простирања кроз логичко коло што за последицу има појаву да се стања улазних сигнала не мењају у истом временском тренутку већ да постоји временско „мимоилажење“ (time skew). На слици 2.19б) приказано је како се правилним распоређивањем и повезивањем тј. балансирањем путања сигнала може избећи ефекат глича а самим тим и минимизирати динамичка потрошња [Pig05]. Ако се елиминацији гличева не посвети дужна пажња динамичка потрошња која је последица транзиционе активности услед гличовања може да чини и до 80% укупне динамичке потрошње [LLW07, LLW08].



Слика 2.19: (а) Појава нежељеног глича, (б) Балансирање путања ради елиминације гличева

С друге стране, у процесу размештања и међусобног повезивања треба избегавати (или свести на минимум) регионе у којима постоје „загушења“ тј. у

којима су везе исувише близу једна другој зато што то може да буде извор шумава или преслушавања између суседних веза.

2.2.1. Утицај размештања и повезивања (PAR) на потрошњу

Из свега горе поменутог може се закључити да фазе размештања и повезивања (PAR фазе) имају велики значај за процес дизајнирања система и да утичу на његове коначне перформансе. Такође, обе фазе, и размештање и повезивање, имају подједнак утицај на карактеристике система. Према томе, за добијање оптималних резултата, обема PAR фазама мора се посветити подједнака пажња тј. обе фазе мора да имају исти приоритет. У супротном може да се деси да предности веома квалитетног размештања намећу нискоквалитетно решење за међусобно повезивање смањујући при томе коначне перформансе система [MN01].

У теорији, PAR проблем спада у групу NP-комлетних проблема и захтева итеративне алгоритме способне да између великог броја потенцијалних решења ефикасно пронађу она решења која су близу оптималног. Неки од широко коришћених таквих алгоритама су *generic arithmetic*, *tabu search* и *simulated annealing* алгоритми [Min01].

Многи аутори истаживали су како током фаза размештања и повезивања минимизирати потрошњу дизајна. У [VKP+10] представљене су две технике за минимизацију потрошње током процеса размештања у FPGA компонентама. Први аспект о коме се дискутовало у раду је функција размештања која за циљ има минимизацију потрошње. Конкретно, описан је модел капацитивности глобалних веза који омогућава редукцију потрошње. Друга техника омогућава редукцију површине и потрошње помоћу оптимизације броја комбинационих и секвенцијалних ћелија. Резултати су квантификовани уз помоћ сета од 119 индустријских бенчмаркова имплементираних на FPGA компоненту са Actel-овом Igl00 архитектуром. Забележена је просечна редукција потрошње од 13% уз просечно побољшање у временским перформансама (кашњењу) од 6.7%.

Чеон (Cheon) са сарадницима представља метод за размештање који је „свестан“ потрошње (труди се да је минимизира) и који истовремено врши груписања регистара у кластере на основу њихове транзиционе активности (како би се редуковала потрошња услед тактовања) и додељивање тежинског фактора везама у зависности од транзиционе активности на њима (за минимизацију потрошње услед транзиционе активности на везама) [CHK+05].

У [GAF+07] инжењери из Xilinx-а разматрају динамичку потрошњу и расположиве CAD технике за минимизацију динамичке потрошње у FPGA компонентама из Virtex-5 фамилије. Техника коју они предлажу састоји се из размештања и повезивања „свесног“ потрошње и нове трансформације која се примењује након повезивања како би се оптимизовала потрошња неких индустријских дизајна. Измерене потрошње на развојним плочама показују да примењене технике редукују потрошњу у просеку за 10%.

Постоје и истраживања на тему редуковања цурећих струја тј. статичке потрошње помоћу различитих PAR техника. Гупта (Gupta) са сарадницима

представља LEAF – нови алат за размештање и повезивање који је „свестан“ статичке потрошње [GDK+07]. Забележена је разлика од чак 190% између PAR алата који су „свесни“ и „несвесни“ статичке потрошње. Сличан алат (тј. алгоритам за размештање и повезивање који минимизира струје цурења) под именом TRAP намењем FPGA компонентама представљен је у [Sal10].

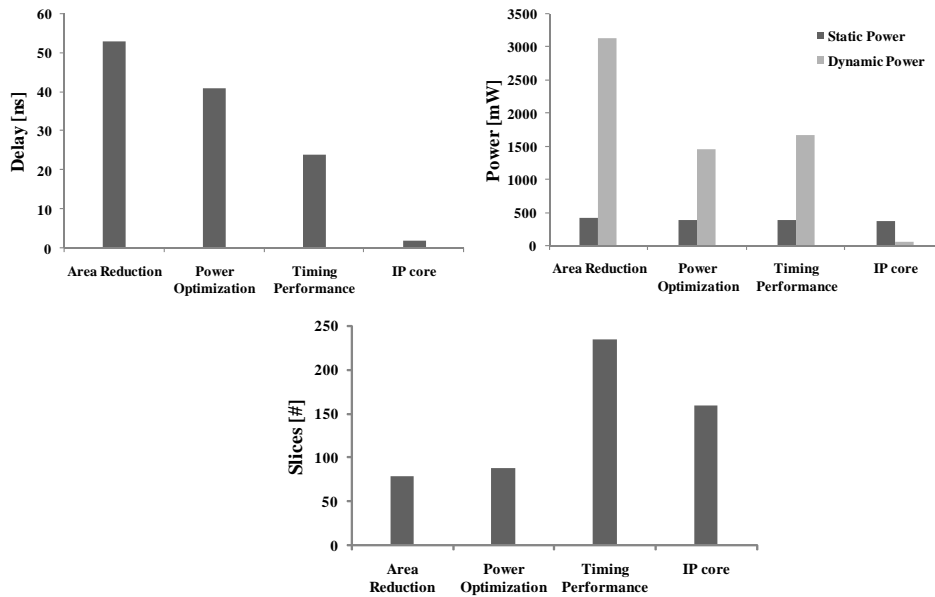
Такође је познато да температурне варијације дуж чипа (термалне неуниформности) представљају претњу по његове перформансе и поузданост. Корелација између укупне потрошње и температурних варијација дуж чипа истраживана је у [NAI11]. Као резултат тога предложене су неке PAR смернице које узимају у обзир поменути корелацију да би ефикасно оптимизовале тоталну потрошњу чипа. Демонстрирано је да се, ако се током размештања и повезивања пажња посвети само минимизацији цурећих струја или само минимизацији максималне температуре на чипу, значајно повећава укупна потрошња чипа. Другим речима, треба вршити размештање и повезивање тако да су утицаји PAR фаза на цуреће струје и максималну температуру на чипу међусобно избалансирани. Експериментални резултати такође показују да смањивање температурних варијација дуж чипа повољно утиче не само на перформансе и поузданост чипа, већ и на минимизацију његове укупне потрошње.

2.2.2. Демонстрација PAR утицаја на перформансе система

Ради демонстрације утицаја размештања и повезивања на свеукупне перформансе система већ описано коло 12-битног бинарног делитеља имплементирано је на FPGA компоненту из Virtex-5 фамилије. Приликом сваке имплементације овог логичког дизајна једини параметар који се мењао био је PAR критеријум. Три различита критеријума по којима се вршило размештање и повезивање били су редуција површине (Area Reduction), оптимизација потрошње (Power Optimization) и минимизација кашњења (Timing Performance). Xilinx-ов PAR алат размештање и повезивање врши у десет итерација и са пет различитих степени „напора“ (1-5). Степен „напора“ одређује количину времена које алат троши приликом налажења што оптималнијег решења. У четвртном случају имплементиран је Xilinx-ов IP core делитељ (који користи исти алгоритам дељења) [Xil06]. На слици 2.20 приказани су резултати имплементације.

Имплементација Xilinx IP core делитеља гарантује најбоље перформансе система у погледу потрошње и брзине рада. Ово се и могло очекивати, имајући у виду да се за IP језгра користе специјални IP блокови у оквиру FPGA компоненте. Ови блокови садрже минималан број транзистора потребан да се имплементира жељена IP логичка функција. Такође, не постоје програмабилне интерконеције што за последицу има минималну могућу капацитивност веза. Са друге стране, три преостала дизајна имплементирана су у FPGA блокове опште намене.

Мењање PAR критеријума за последицу има очигледну варијацију у перформансама система. Тако, флукутације у брзини и динамичкој потрошњи могу ићи и до 55%, док величина дизајна варира 65% (у односу на најгори случај). Статичка потрошња незнатно се мења (око 6%).



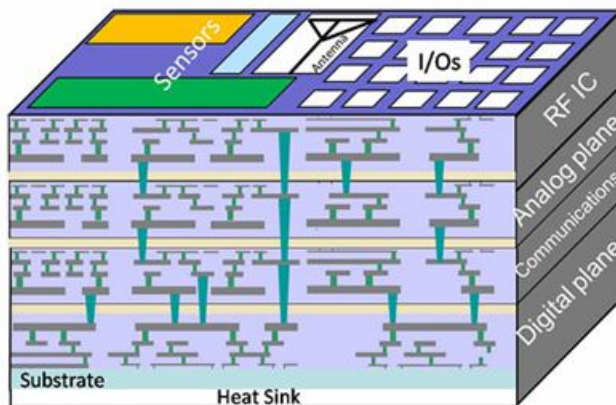
Слика 2.20: Утицај PAR критеријума на перформансе дизајна

2.2.3. 3D интегрисана кола

Перформансе VLSI интегрисаних кола произведених у нанометарским технологијама (deep-submicron) све више и више ограничавају се од стране интерконекција на чипу које полако постају камен спотицања даљег напретка. Једноставно, решавање проблема интерконекција (смањивање кашњења и потрошње на везама), као и проблема који ограничавају максимални број улазних и излазних сигнала у колу у великом је раскорак (спорије напредује) у односу на брзину побољшавања карактеристика транзистора. Последица овога су лимитиране перформансе интегрисаних кола. Увођење „треће димензије“ у интегрисано коло тј. примена концепта слагања слојева активних компоненти (слика 2.21) редуковало би, у односу на 2D интегрисана кола, дужину интерконекција, побољшало ефикасност процесирања податка по јединици запремине а такође би и олакшало могућност интеграције хетерогених система (од којих се углавном састоји дизајн система на чипу – System-on-Chip (SoC) дизајн). Предности 3D интегрисаних кола препознате су у Intel-у чији су инжењери 2004. године произвели и представили 3D Pentium 4 процесор [BNW+04]. Разбијање великих блокова као и блокова који су велики потрошачи и њихово пажљиво реаранжирање које је омогућено 3D дизајном ограничило је број врућих тачака у дизајну. 3D структура процесора за 15% је побољшала његове перформансе, а за исти проценат је смањена и потрошња (као последица смањивања дужине интерконекција и елиминације репетитора сигнала).

У [PF09] представљена је студија о 3D интегрисаним колима. Дате су технологије њихове производње, производње паковања (кућишта) за 3D интегрисана кола, технике физичког дизајна и архитектуре 3D кола, технике за минимизацију врућих тачака тј. термалне дисипације 3D интегрисаних кола као и дизајнирање њихове дистрибуционе мреже такта.

Потрошња електричне енергије у 3D интегрисаним колима, алати за процену заузећа површине, перформанси, потрошње, прављење температурних мапа 3D интегрисаних кола, као и технике размештања и повезивања (PAR) и њихов утицај на потрошњу описани су у [BSK+01, DCR03, AFG+05, SS08, Bor09].



Слика 2.21: 3D интегрисано коло

2.3. Оптимизација на нивоу архитектуре система

Једна иста логичка функција (произвољна аритметичка функција, банка филтара, структура микропроцесора, стаза података итд.) може се имплементирати употребом различитих архитектура које се међусобно разликују у степену паралелизма или проточности, дужини критичне путање, броју транзистора итд. Под претпоставком да све те различите архитектуре нису исти потрошачи електричне енергије, закључује се да је у скупу могућих архитектура, у погледу потрошње електричне енергије, једна од њих оптимална тј. да има минималну потрошњу. Имајући у виду сталан тренд пораста статичке потрошње архитектуре дизајна потенцијални кандидати за минималну укупну потрошњу би свакако требало да је не занемаре. Са јако „цурећим“ транзисторима на располагању, логично је да треба избегавати неактивне гејтове (тј. гејтове који ће врло ретко да мењају своје стање). Овакви гејтови били би у стању мировања (idle) током већег дела времена функционисања система и не би активно доприносили логичкој функцији док би са друге стране знатно повећавали статичку потрошњу. Имајући то у виду, за задату логичку функцију, архитектура са малим бројем врло активних гејтова могла би да има предност у односу на архитектуру са великим бројем мало активних гејтова.

На пример, логичка функција која захтева 10000 гејтова за своју имплементацију и испољава активност од 1% има 100 транзиција у току једног тактног интервала. Ако даље претпоставимо да иста таква логичка функција може да се имплементира са 1000 гејтова и да такође има 100 транзиција током једне периоде такта закључујемо да ће активност у другој архитектури нарасти на 10%. Иако ће обе архитектуре да имају исту динамичку потрошњу (јер имају исти број транзиција по периоди такта), статичка потрошња друге архитектуре смањиће се 10 пута тј. за онолико за колико је број гејтова у другој архитектури мањи у односу на прву.

Заиста, смањени број гејтова са задржаним (истим) бројем транзиционих прелаза свакако за последицу има повећану активност (дефинисану као однос броја гејтова који мењају своје стање и броја укупних гејтова). Ово је у супротности са до сада познатим методологијама дизајнирања које су нису узимале у обзир статичку потрошњу и које су тежиле да смање активност како би тиме минимизирале динамичку потрошњу. Треба рећи и то да технике за минимизацију потрошње на овом нивоу не мењају технолошке параметре дизајна (напон напајања и прага, димензије транзистора, материјале за израду) и да не утичу на размештање и повезивање истог. На потрошњу дизајна утиче се искључиво преко његове архитектуре (паралелизам и проточност различитих нивоа, RNS и LNS аритметика итд.).

2.3.1. Методологија селекције оптималне архитектуре

Шустер (Schuster) је са сарадницима представио методологију која омогућава, да се из скупа различитих потенцијалних архитектура дизајна изабере она која ће имати минималну укупну (статичку плус динамичку) потрошњу [SPN+05]. У основи ове методологије налази се следећа једначина за рачунања укупне потрошње дизајна:

$$P_{tot} = P_{dyn} + P_{stat} = a \cdot N \cdot C \cdot V_{dd}^2 \cdot f + N \cdot V_{dd} \cdot I_o \cdot \exp\left(\frac{-(V_{th} - \eta \cdot V_{dd})}{nU_t}\right) \quad (2.3)$$

где је a : просечна активност (број гејтова који током једног тактног интервала мења своје стање (укључујући и гличеве) у односу на укупан број гејтова); N : укупан број ћелија дизајна; C : еквивалентна капацитивност ћелије; I_o : просечне струје цурења за $V_{GS}=V_{th}$; η : DIBL коефицијент; n : нагиб у режиму слабе инверзије; U_t : термални напон ($=kT/q$). Ради упрошћења модела статичка потрошња одређена је само струјом цурења у каналу. Статичку потрошњу могуће је детаљније моделовати тј. проширити је и осталим цурећим струјама. Ово проширење неће утицати на методологију селекције. Крајњи циљ методологије биће идаље исти: за фиксне вредности напона напајања и напона прага, у групи могућих архитектура дизајна, одредити архитектуру са најмањом укупном потрошњом електричне енергије. У пракси је полазна претпоставка методологије о фиксним вредностима напона напајања и прага најчешће исправна, будући да дизајнери врло ретко могу да бирају оптималне вредности напона којима ће напајати поједине логичке блокове. Према томе, све архитектуре из скупа – потенцијални кандидати за најмању потрошњу имају исте напоне напајања и прагова, као и радну фреквенцију f . Једино по чему се архитектуре међусобно разликују су различите вредности за активност a као и укупан број ћелија N . Две архитектуре су, дакле, одређене својим вредностима за a_1 и N_1 и a_2 и N_2 , респективно. Њихова укупна потрошња електричне енергије може да се упореди на следећи начин:

$$\begin{aligned} a_1 \cdot N_1 \cdot C \cdot V_{dd}^2 \cdot f + N_1 \cdot V_{dd} \cdot I_o \cdot \exp\left(\frac{-(V_{th} - \eta \cdot V_{dd})}{nU_t}\right) < ? \\ a_2 \cdot N_2 \cdot C \cdot V_{dd}^2 \cdot f + N_2 \cdot V_{dd} \cdot I_o \cdot \exp\left(\frac{-(V_{th} - \eta \cdot V_{dd})}{nU_t}\right) \end{aligned} \quad (2.4)$$

Неједначина 2.4 исправна је у случају да прва архитектура конзумира мање енергије у односу на другу. Ову неједначину могуће је преформулисати на следећи начин:

$$(N_1 - N_2) < ? - (a_1 \cdot N_1 - a_2 \cdot N_2) \frac{C \cdot V_{dd} \cdot f}{I_0 \cdot \exp\left(\frac{-(V_{th} - \eta \cdot V_{dd})}{n \cdot U_t}\right)} \quad (2.5)$$

Након тога, дефинишући разлику у броју ћелија као $\Delta N = N_1 - N_2$ и разлику у активности две архитектуре као $\Delta Tr = a_1 \cdot N_1 - a_2 \cdot N_2$, поређење између две архитектуре може да се сведе на следеће:

$$\Delta N < ? - \Delta Tr \frac{C \cdot f}{I_0} \frac{V_{dd}}{\exp\left(\frac{-(V_{th} - \eta \cdot V_{dd})}{n \cdot U_t}\right)} \quad (2.6)$$

$$\Delta N < ? - \Delta Tr \cdot R(V_{dd}, V_{th}, f)$$

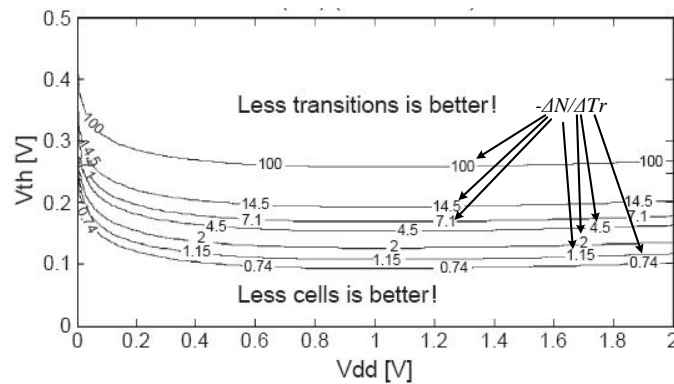
Функција $R(V_{dd}, V_{th}, f)$ у неједначини 2.6 увек је позитивна и зависи од V_{dd} , V_{th} , f као и од неких технолошких параметара који су дизајнеру наметнути и које он не може да мења. Неједначина 2.6 показује да поређење укупне потрошње две различите архитектуре једног дизајна зависи од разлике у броју њихових ћелија (ΔN) као и од разлике у њиховим активностима (ΔTr). Овај закључак се прилично разликује од конвенционалних приступа код којих се пореди само активност архитектура будући да се у обзир узима само динамичка потрошња енергије.

Ако оба дизајна испољавају исту активност и имају исти број ћелија ($\Delta Tr = 0$ и $\Delta N = 0$) логично је очекивати да ће имати и исту потрошњу. Архитектура са више ћелија и већом активношћу ће свакако увек имати и већу потрошњу. У том случају неједначина 2.6 постаје тривијална тј. независна од R . Насупрот томе, ако једна архитектура има више ћелија и мању активност у односу на другу ($\Delta N > 0$ и $\Delta Tr < 0$), и обрнуто ($\Delta N < 0$ и $\Delta Tr > 0$), избор архитектуре која троши мање енергије постаје комплекснији и зависи од R . То значи да ће радни услови тј. V_{dd} , V_{th} , f као и неки технолошки параметри одредити која ће од две архитектуре бити мањи потрошач електричне енергије. Сви могући случајеви међусобног поређења две архитектуре сумирани су у табели 2.2.

Табела 2.2: Међусобно поређење две архитектуре једног дизајна које имају одређену разлику у броју ћелија ΔN и у транзиционој активности ΔTr . Архитектура назначена у табели је она са мањом укупном потрошњом електричне енергије

	$\Delta Tr > 0$	$\Delta Tr = 0$	$\Delta Tr < 0$
$\Delta N > 0$	Архитектура 2	Архитектура 2	Зависи од R
$\Delta N = 0$	Архитектура 2	Иста потрошња	Архитектура 1
$\Delta N < 0$	Зависи од R	Архитектура 1	Архитектура 1

Боље разумевање улоге коју има параметар R на селекцију архитектуре читалац може стећи анализом слике 2.22 на којој су приказане линије једнаке потрошње два дизајна у $V_{dd} - V_{th}$ равни. Свакој линији на графику одговара једна вредност параметра $-\Delta N / \Delta Tr$ а линија представља скуп тачака у којима две архитектуре имају исту потрошњу (потрошња ће варирати у зависности од положаја тачке тј. од вредности V_{dd} и V_{th} , али ће бити иста за обе архитектуре). На пример, ако две архитектуре које раде на $V_{dd} = 1V$ и $V_{th} = 0.27V$ имају параметар $-\Delta N / \Delta Tr = 100$ њихова потрошња електричне енергије биће идентична.



Слика 2.22: Линије константне потрошње у V_{dd} - V_{th} равни за $f=31.25\text{MHz}$ и дизајн пројектован у UMC $0.18\mu\text{m}$ технологији. Пар архитектура који се посматра одговара једној кривој на графику тј. једној вредности параметра $-\Delta N/\Delta Tr$

Иначе, када су дизајнеру наметнута таква ограничења у погледу V_{dd} и V_{th} да је радна тачка изнад линије једнаке потрошње (нпр. $V_{dd}=1\text{V}$ и $V_{th}=0.4\text{V}$), оно коло које испољава мање транзиција конзумираће и мање енергије. Ова констатацију треба поткрепити чињеницом да тачке које су изнад линије једнаке потрошње углавном имају већу вредност напона прага од тачки на линији једнаке потрошње. Повећањем напона прага смањује се статичка потрошња што утиче на то да динамичка компонента потрошње постане доминантна. У том случају, јасно је да ће дизајн који испољава мање транзиционе активности имати мању динамичку, а самим тим и мању укупну потрошњу. Насупрот овоме, кад је радна тачка испод линије једнаке потрошње (нпр. $V_{dd}=1\text{V}$ и $V_{th}=0.2\text{V}$) дизајн са мање ћелија биће мањи потрошач. У овом случају напон прага смањен је у односу на тачке које су на линији једнаке потрошње. Смањена вредност напона прага повећава статичку потрошњу која постаје доминантна. Мањи број ћелија сада значи и мању статичку а самим тим и укупну потрошњу.

Управо описана методологија међусобног поређења две архитектуре може се итеративно применити на већи скуп потенцијалних архитектура. На тај начин, понављањем међусобног поређења две архитектуре, могуће је елиминисати оне најгоре и брзо конвергирати ка најбољој, тј. оној са најмањом потрошњом. Битно је приметити и то да је архитектура са најмањом потрошњом одређена и технолошким параметрима на које дизајнер не може да утиче. Променом технолошких параметара може да се деси да нека друга архитектура из скупа буде најмањи потрошач.

Уопштено, процес одређивања архитектуре са најмањом потрошњом састоји се у следећем:

1. **Временска ограничења:** За задате параметре V_{dd} , V_{th} и f елиминишу се архитектуре које су исувише „споре“ тј. не испуњавају унапред задата ограничења у погледу брзине рада.
2. **Поређење пара архитектура:** Поређење две архитектуре врши се уз помоћ параметра $-\Delta N/\Delta Tr$. Ако је његова вредност негативна, која ће архитектура од две које се пореде имати мању потрошњу зависи од броја ћелија у архитектурама као и од њихове транзиционе активности (види табелу 2.2). У другом случају, $-\Delta N/\Delta Tr > 0$, избор архитектуре биће одређен параметром R тј. положајем радне тачке у односу на линију једнаке потрошње.

3. Понови корак 2 за све преостале архитектуре: Поступак одређивања архитектуре-победника могао би се убрзати и процесом елиминације тривијалних случајева тј. архитектура које очигледно неће конзумирати најмање енергије јер у односу на друге архитектуре испољавају већу активност и у исто време имају већи број ћелија. Тиме би се смањио број међусобних поређења тј. брже би се конвергирало ка архитектури која за задата технолошка ограничења (V_{dd} , V_{th} и f) троши најмање енергије.

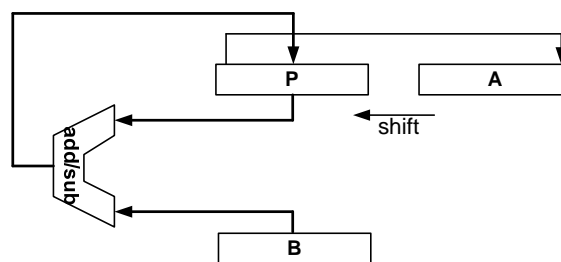
У [SPN+05] аутори су описану методологију применили на скуп од 11 архитектура које су обављале множење два 16-битна броја. Архитектуре су се међусобно разликовале како по структури тако и по степенима паралелизма и проточности. Методологија је на исти сет архитектура примењена два пута. Једино по чему се прва примена разликовала од друге био је положај радне тачке у V_{dd} - V_{th} равни. Различити технолошки параметри утицали су на то да се за оптималну архитектуру у погледу потрошње у две одвојене примене методологије добију две различите архитектуре.

2.3.2. Демонстрација методологије селекције оптималне архитектуре

Технике проточности (pipelining) и паралелизма различитих нивоа примењене су на *Radix-2 non-restoring* и *Sequential (restoring)* архитектуре 12-битних бинарних делитеља како би се на скуп од 8 различитих архитектура применила методологија селекције оптималне архитектуре у погледу потрошње електричне енергије. *Radix-2 non-restoring* алгоритам дељења већ је описан и представљен на слици 2.16, док је алгоритам секвенцијалног (*Sequential*) делитеља целобројних означених бројева представљених у другом комплементу приказан на слици 2.23.

Процес секвенцијалног дељења може се описати на следећи начин: Постави дељеник X у регистар A , делилац Y у регистар B , ресетуј регистар P . M пута обави следеће кораке дељења (M је број битова дељеника X):

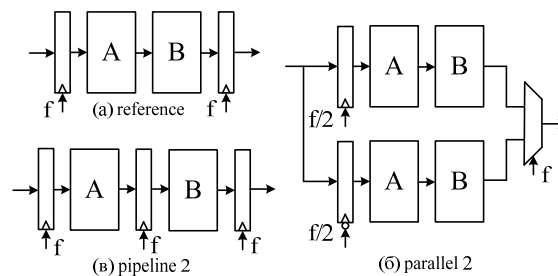
- Шифтуј регистарски пар (P,A) за једно месту улево
- Одузми садржај регистра B од садржаја регистра P , смести резултат у P
- Ако је резултат одузимања негативан постави бит најмање тежине (LSB) у регистру A на 0, иначе на 1
- Ако је резултат одузимања негативан врати стари садржај регистра P (додавањем садржаја регистра B)



Слика 2.23: *Sequential* алгоритам бинарног дељења

Након M оваких корака регистар A садржаће количник дељења док ће садржај регистра P бити једнак остатку дељења. Више о поменутим архитектурама бинарног дељења и резултатима примена паралелизма и проточности различитих степена са становишта величине дизајна и брзине рада може се видети у [Ј10].

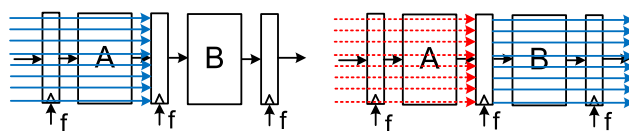
Брзина рада неког дизајна одређена је његовом критичном путањом тј. путањом између два чвора у колу која су међусобно повезана. Путању зовемо критичном зато што она има најдуже пропационо кашњење сигнала, дуже од било које друге путање у дизајну. Ако са t_{cp} означимо време простирања сигнала кроз критичну путању, максимални „проток информација“ кроз дизајн тј. његова брзина рада приближно је једнака $f_{max} \approx 1/t_{cp}$. Паралелизам и проточност су опште познате технике намењене пре свега за повећање „протока информација“ тј. брзине рада дизајна. На слици 2.24 приказан је референтни дизајн заједно са својом паралелизованом и проточном верзијом степена 2.



Слика 2.24: Пример микро-архитектуре: (а) Референтни дизајн, (б) Паралелизам степена 2, (в) Проточност степена 2

Паралелизам заправо представља просту репродукцију дизајна. Број репродукција дизајна зависи од степена паралелизма. Свака репродукција дизајна процесира део улазних података. На тај начин се може или повећати „протока информација“ (брзина рада) или извршити „растеређење“ дизајна (задржавањем брзине рада референтног дизајна). На слици 2.24б) половина улазних података процесира се једном репродукцијом дизајна док се друга половина обрађује другом репродукцијом. Након обраде података, уз помоћ мултиплексера, резултати обраде се наизменично пропуштају на излаз. Увођењем паралелизма не мења се критична путања дизајна. Повећања протока информација последица је чињенице да се функција дизајна обавља паралелно и независно на две (или више, у зависности од степена паралелизма) репродукције дизајна – $f_{max} \approx k/t_{cp}$, где је k степен паралелизма.

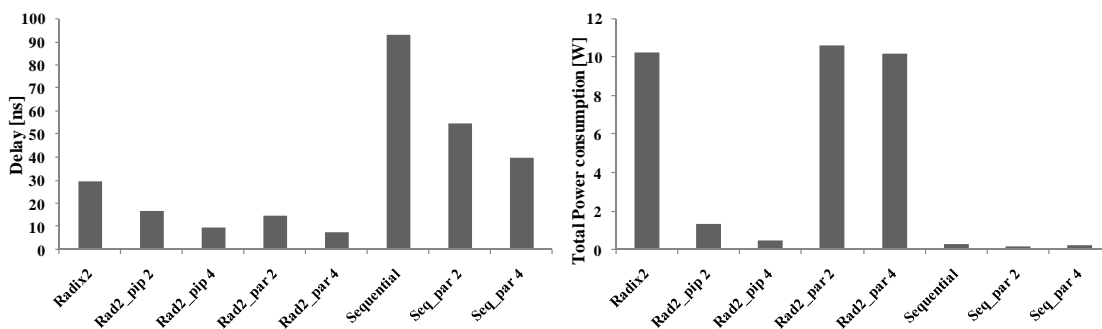
Техника проточности састоји се у томе да се на одговарајућим местима у дизајну поставе регистри који би прихватили талас улазних података који се креће ка излазу кола. Након што се овај талас података упише у регистре могуће је прихватити нови талас улазних података. Ово је илустровано на слици 2.25.



Слика 2.25: Илустрација технике проточности степена 2

Степен проточности одређен је бројем места у дизајну на којима се врши регистровања таласа података. Треба рећи и то да технику проточности није могуће применити на сваки дизајн. Применом технике проточности смањује се критична путања у дизајну што за последицу има повећање протока информација тј. брзине рада. Поред тога што незнатно повећава површину дизајна (због додавања регистара) проточни дизајн има и већу латенцију (број тактних циклуса од тренутка кад се поставе улазни подаци до тренутка кад се добију валидни резултати) у односу на референтни.

Свих 8 различитих архитектура бинарног дељења имплементирани су на FPGA компоненту из Virtex-4 фамилије. Такође, уз помоћ Cadence алата и AMIS 0.35 μ m технологије, извршена је HDL (Hardware Description Language) синтеза и ASIC физичка имплементација архитектура. Технолошки параметри свих 8 архитектура (V_{dd} , V_{th} , f) били су исти. На референтне архитектуре примењивана је техника паралелизма степена 2 и 4, а у случају *Radix-2 non-restoring* алгоритма дељења и техника проточности степена 2 и 4. Слика 2.26 и Табела 2.3 представљају резултате FPGA и ASIC имплементације, респективно.



Слика 2.26: Кашњење и укупна потрошња сета од 8 архитектура бинарног делитеља имплементираних на Virtex-4 FPGA

Кад се обавља паралелизација број ћелија у дизајну се отприлике дуплира ($\Delta N < 0$, под претпоставком да референтни дизајн има N_1 ћелија а паралелизовани N_2). С друге стране, активност паралелизоване архитектуре се смањује нешто мање него дупло. Како је повећање у броју ћелија углавном доминантније је у односу на смањење активности ($\Delta Tr < 0$) паралелизована верзија ће, под истим радним условима, увек трошити више (или приближно исто) енергије у односу на референтну архитектуру. Ови закључци потврђени су и на слици 2.26 и у табелама 2.2 и 2.3. Ипак, у случају да референтна архитектура не задовољава унапред задата временска ограничења паралелизација може послужити као ефикасно средство за повећање брзине рада (види слику 2.26).

Табела 2.3: Резултати имплементације бинарних делитеља на ASIC чип AMIS 0.35 μ m технологије

	Cell Area [μ m ²]	Cells	Average Cell Power [mW]
Radix2	1374	515	20.589
Radix2_pipeline 2	1615	536	9.289
Radix2_pipeline 4	2054	639	6.072
Radix2_parallel 2	2525	922	20.093
Radix2_parallel 4	5060	1865	20.082
Sequential	857	342	0.382
Seq_parallel 2	1563	666	0.709
Seq_parallel 4	3136	1353	1.361

Примена технике проточности за последицу има незнатно повећање броја ћелија дизајна (услед употребе регистара, $\Delta N < 0$) док је активност доста редукована, углавном због елиминације гличева ($\Delta Tr > 0$). Другим речима, смањење активности доминира у односу на повећање броја ћелија. Из табеле 2.2 може се видети да је, у случају примене проточности, архитектура са мањом потрошњом одређена параметром R (радним условима и технолошким параметрима). У случају наших имплементација техника проточности редуковала је потрошњу дизајна. Поред потрошње, редукована је и критична путања у дизајну што за последицу има повећање брзине рада.

Резултати имплементације такође показују да архитектура секвенцијалног делитеља конзумира много мање енергије. Ова уштеда у потрошњи плаћена је великим повећањем кашњења тј. смањењем брзине рада. Такође, може се закључити да су способности програмабилности и флексибилности FPGA компоненте плаћене повећаном потрошњом електричне енергије у односу на ASIC интегрисано коло. FPGA блокови за конфигурацију (програмабилна мрежа за повезивање логичких блокова) и конфигурациона меморија знатно повећавају површину силицијума потребну за имплементацију што за последицу има повећане дужине веза за повезивање као и повећану капацитивност интерконекија а самим тим и повећану укупну потрошњу.

2.3.3. Остале минимизационе технике на нивоу архитектуре система

Поред паралелизма и проточности различитих степена могуће је вршити и неке друге модификације архитектуре чији је циљ повећање ефикасности обраде података, смањење потрошње електричне енергије итд. У наставку ће у кратким цртама бити описане технике „рачунања унапред“, прерасподеле регистара, као и употреба RNS и LNS аритметике.

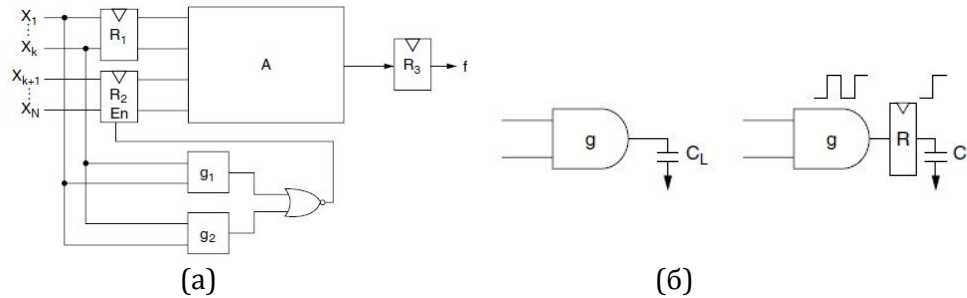
Техника рачунања унапред састоји се у селективном рачунању унапред логичких стања на излазу кола (један тактни циклус пре него што су потребна). Овако добијена логичка стања се након тога користе чиме се редукује интерна транзициона активност комбинационе логике у следећем тактном интервалу [AMD+94].

Један прост пример ове технике приказан је на слици 2.27а) на којој су улази секвенцијалног блока A подељени у два сета и воде се на улазе регистара R_1 и R_2 , док се излаз из блока A води на улаз регистра R_3 . Булове функције g_1 и g_2 имају улогу предиктор функција и то на следећи начин: ако је $g_1=1$ онда је $f=1$, ако је $g_2=1$ онда је $f=0$. Другим речима, ако је барем једна од Булових функција g_1 и g_2 у стању логичке '1' вредност излазне функције f је у потпуности одређена. Како функције g_1 и g_2 зависе само од дела улазних променљивих блока A (X_1, \dots, X_k), преостале променљиве (X_1, \dots, X_k) могуће је замрзнути помоћу En сигнала на регистру R_2 .

Према томе, ако је једна од предикторских функција, g_1 или g_2 у стању логичке '1' за време тактног циклуса T , сигнал дозволе регистра R_2 (En) поставља се у стање логичке '0'. Као последица тога, излаз регистра R_2 се за време тактног циклуса $T+1$ не мења док се излазна функција f ипак поставља у исправно логичко стање (због тога што се мења стање на излазу регистра R_1). Крајњи резултат је смањена транзициона активност у блоку A а самим тим и значајне уштеде у

потрошњи електричне енергије. Због тога што функције g_1 и g_2 повећавају површину интегрисаног кола а такође и троше додатну енергију потребно је извршити одговарајућу анализу компромиса тј. утврдити да ли је уштеда у потрошњи електричне енергије у блоку A већа од потрошње енергије предикторских функција.

Још један метод који редукује потрошњу у проточним секвенцијалним колима представљен је у [MDG93]. Састоји се у прерасподели регистра (флип-флопова) унутар структуре интегрисаног кола што за последицу може да има минимизацију површине, кашњења или потрошње електричне енергије.



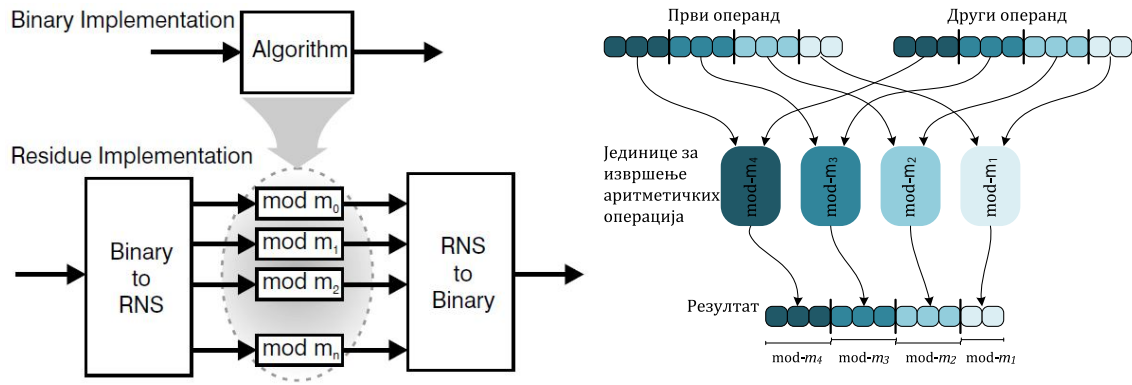
Слика 2.27: (а) Структура за рачунање унапед код секвенцијалних кола, (б) Прерасподела регистра за минимизацију потрошње

Идеја овог метода је да се у чворовима интегрисаног кола који имају велики глич или капацитивност поставе флип-флопови (због њихове особине да, када се тактују, њихов излаз може да направи највише једну транзицију тј. промену стања). На тај начин се спречава пропацирање гличева и редукација укупне транзиционе активности тј. потрошње електричне енергије (види слику 2.27б). Ипак, треба бити опрезан са постављањем оваквих регистра из разлога што су и они сами, заједно са линијама такта, значајни потрошачи, а уз то и повећавају површину интегрисаног кола.

Суштина RNS (Residue Number System) аритметике (чије су основе постављене у 4. веку нове ере) је да се, као што је приказано на слици 2.28, гломазни операнди поделе на мање и да се над њима примени паралелна обрада. Тиме ће се добити на брзини обраде података, аритметичке операције ће се поједноставити (пошто су операције међусобно независне не постоји пропација преноса и позајмица између канала) а уштедеће се и потрошња електричне енергије. RNS аритметика најчешће се примењује на DSP архитектуре код којих доминира сабирање и множење као нпр. код дигиталних филтера и конволуција [CN04]. Након поделе операнада и обраде података резултате обраде потребно је објединити тј. извршити конверзију из RNS репрезентације резултата у његову бинарну репрезентацију. Ова конверзија представља уско грло RNS аритметике и у многоме одређује особине (брзину рада, потрошњу итд.) RNS архитектуре. Три основна технике конверзије RNS бројева у бинарне су кинеска теорема остатака (Chinese Remainder Theorem – CRT) [GC08], нова кинеска теорема остатака [MND+10] и mixed-radix конверзија (Mixed Radix Conversion – MRC) [SJ11].

У студији о потрошњи електричне енергије RNS архитектуре FIR (Finite Impulse Response) филтра аутори у [FP97] закључују да RNS архитектура може смањити потрошњу будући да смањује цену хардвера, транзициону активност и напон напајања. Ради демонстрације утицаја RNS архитектуре на потрошњу

филтра аутори су показали да основна архитектура FIR филтра са 16-битним коефицијентима, 32-битним динамичким опсегом и аритметиком у другом комплементу, на радној фреквенцији од 50MHz у просеку троши 26.2mW. Са друге стране RNS еквивалент FIR филтра троши свега 3.8mW. Уштеда енергије постаје још већа са повећавањем реда филтра. Још неке студије о утицају RNS архитектуре на потрошњу дизајна могу се наћи у [Chr98, GVC10].



Слика 2.28: Блок шема система који обавља аритметичке операције над RNS бројевима

Принцип обраде података код LNS (Logarithmic Number System) архитектура, као и код њене комплексне варијанте, CLNS, у основи је сличан RNS обради података. Нешто више о LNS и CLNS аритметици може се наћи у [Lew99, FLT+09].

Глава 3

Метода верификације модела за процену потрошње

Имајући у виду да је, према слици 1.5, највећу уштеду потрошње могуће остварити у почетним фазама пројектовања тј. на највишим нивоима апстракције дизајна, и да, у исто време, процена потрошње на овим нивоима није временски захтевна, у наставку дисертације биће описана методологија естимације динамичке потрошње на високом нивоу апстракције. Методологија је примењена на DSP дизајне имплементирание на FPGA компоненте будићи да оне, због своје флексибилности и програмабилности, представљају неку врсту компромиса између спорих, гломазних и „гладних“ енергије микропроцесора и ASIC чипова специјализованих само за одређену функцију. Нешто више о самој структури FPGA компоненти биће речи у наставку дисертације.

Тачност модела процене потрошње јако је зависна од тачности метода који се користе за њихову карактеризацију тј. од референтних вредности са којима се процењена вредност потрошње пореди. За комерцијалне FPGA компоненте постоји неколико алата за процену потрошње на ниском нивоу. Базирају се на симулацијама дизајна чија се процена потрошње врши, а неки од најчешће коришћених су XPower из Xilinx-а и Altera Power Play. Ови алати дају детаљан извештај о потрошњи дизајна на основу информација о капацитивностима, искоришћењу FPGA компоненте и транзиционој активности дизајна за задате сигнале на улазу. Ипак, у [LLC+05, ESJ06, JJC10] примећене су велике грешке између измерене потрошње и процењене потрошње коју су дали ови алати. Додатни проблеми јављају се код комплекснијих дизајна код којих је потребно моделовати много сигнала. У том случају ови алати захтевају доста радне меморије и времена потребног да се изврши процена. Имајући ово у виду, одлучено је да се наши модели за процену динамичке потрошње карактеришу тј. верификују путем мерења потрошње дизајна који се имплементира на развојној плочи.

Услед различите капацитивности интерконекција и логике дизајна (конфигурабилних логичких блокова (CLBs) FPGA компоненте), до процењене вредности потрошње ових елемената долази се преко различитих модела. Према томе, потребно је наћи начина да се измерена вредност потрошње рашчлани на динамичку потрошњу интерконекција (глобалних) и динамичку потрошњу логике. Даље, да би процењене вредности потрошње биле што тачније, од динамичке потрошње логике неопходно је одузети потрошњу улазних бафера. Имајући ово у виду мерну апаратуру неопходно је пажљиво дизајнирати како би се резултати мерења потрошње могли накнадно обрађивати.

У наставку ове главе најпре је дат кратак опис FPGA програмабилних компоненти. Такође је описана и архитектура Virtex II Pro компоненте која се налази на XUP развојној плочи. Након тога, описана је методологија мерења потрошње коришћена у овој тези. Ова методологија развијена је на Политехничком Универзитету у Мадриду и састоји се од мерне поставке као и од неких алата развијених са циљем да скупе информације о имплементираним (размештеном и повезаном) дизајну и да се екстракују капацитивности електричних веза за повезивање. На крају је представљен и XPower алат чије су нам процењене вредности потрошње такође служиле као референца.

3.1. Структура FPGA компоненти

FPGA компонента може се описати као дводимензионални низ логичких блокова и флип-флопова који су међусобно повезани електрично програмабилним везама. За разлику од интегрисаних кола специфичне намене (ASIC) која могу да током свог радног века врше само једну одређену функцију, FPGA компоненте могуће је репрограмирати за врло кратко време (реда милесекунде [WBG+06]). Ипак, флексибилност FPGA компоненте плаћена је повећањем сложености њене структуре. Треба рећи и то се последњих година у FPGA структуру додају неки нови ресурси (уграђене меморије, множачи, процесори итд.) како би се проширила област њихове примене.

У зависности од технолошког процеса производње, постоје два основна типа FPGA компоненти. Први тип су такозване SRAM-базиране FPGA компоненте код којих је програмабилност постигнута преко специјалних несталних (volatile) SRAM меморијских ћелија које активирају или деактивирају програмабилну логику. Највећи произвођачи овог тима FPGA компоненти су Xilinx, Altera, Atmel и Lattice Semiconductor Corporation. Други тип су *antifuse* и флеш-базиране FPGA компоненте које се програмирају стварањем сталних проводних електричних путања као у случају anti-осигурача (*antifuse*) или преко постојаних (non-volatile) флеш ћелија које делују на програмабилну логику. Најпознатији произвођачи су Actel и QuickLogic corporation [KTR07].

Ипак, водећи глобални произвођачи програмабилне логике су Xilinx и Altera који покривају 80% укупног тржишта. Због тога су се за мерење потрошње користиле компоненте ових произвођача.

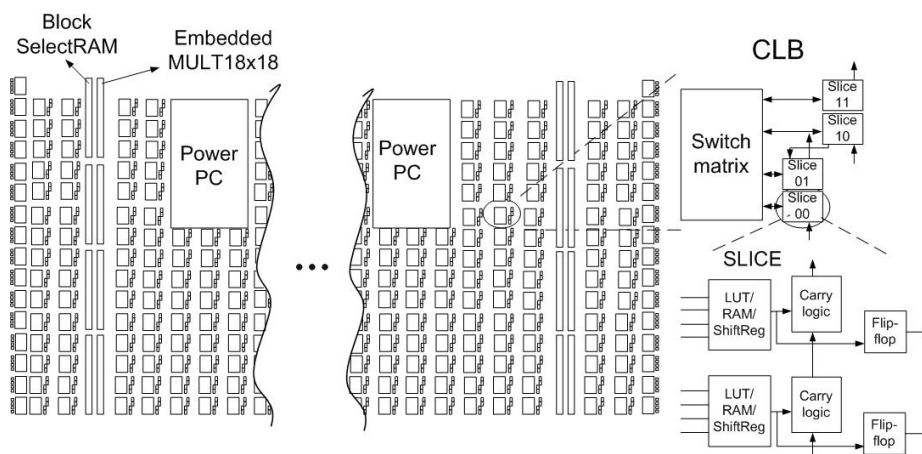
Прва значајна имплементација апликације која је користила операнде двоструке тачности са покретним зарезом десила се 2002 године, са појавом Xilinx-ових Virtex II Pro компоненти које су садржале до 100.000 логичких ћелија, уграђене 18x18 множаче и PowerPC RISC процесорске блокове. Ово је један од

разлога због кога смо за циљну FPGA компоненту изабрали баш Virtex II Pro XC2VP30-7 и XUP (Xilinx University Program) развојну плочу. Virtex-4 фамилија FPGA компоненти појавила се на тржишту 2004. године била је надоградња Virtex II Pro фамилије, са 200.000 логичких ћелија и уграђених DSP блокова способних за обављање операција множења и акумулирања (MAC). На жалост, уграђени множачи (18x18) нису били добро прилагођени за операције са покретним зарезом. Да би се разрешио овај проблем 2008. године појавиле су се FPGA компоненте из Virtex-5 фамилије. Садржале су велики број 25x18 уграђених множача.

Иако је у овој дисертацији коришћена FPGA компонента из Virtex II Pro фамилије описана методологија лако је применљива и на FPGA компоненте из најновијих фамилија, будући да се све оне заснивају на Virtex II Pro архитектури.

3.1.1. Virtex II Pro архитектура

Virtex II Pro архитектура приказана је на слици 3.1. Сваки конфигурабилни логички блок (CLB) из дводимензионалног низа садржи четири “парчета” (slice) и повезан је са мрежом за глобално повезивање конфигурабилних блокова (switch matrix). Свако парче (слајс) садржи по две look-up табеле (LUTs) које служе за генерисање било које логичке функције која може да зависи највише од четири променљивих. Ове табеле могу да се конфигуришу и као померачки регистри (ShiftReg) или RAM меморијски блокови. Поред LUT-ова, слајс садржи и флип-флопове, логичке гејтове и мултиплексере.

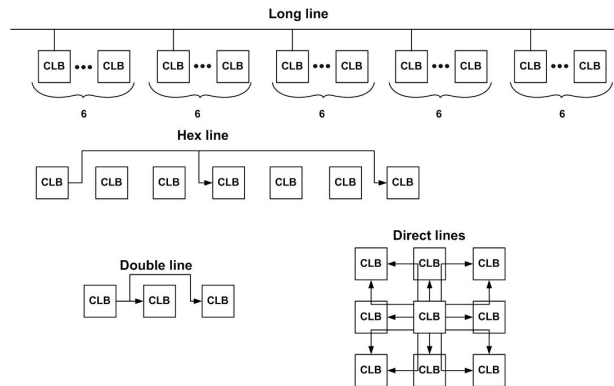


Слика 3.1: Virtex II Pro архитектура

Virtex II Pro архитектура садржи и Dual-port RAM меморијске блокове (Block SelectRAM) капацитета 18Kb чију је дубину и ширину могуће мењати (од 16K x 1b до 512 x 32 b). Уграђени 18x18 множачи оптимизовани су за операције акумулирања и множење (MAC) као и за примену у дигиталним филтрима. RAM меморијски блокови и уграђени множачи такође су преко матрица прекидача (switch matrix) повезани на мрежу за глобално повезивање. PowerPC блокови су IBM RISC а њихов број може да варира између 0 и 2.

Сви блокови у Virtex II Pro архитектури (улази/излази, конфигурабилни логички блокови, RAM меморијски блокови и множачи) користе исту мрежу за

међусобно повезивање. Интерконекције су бафероване што их чини независним од fanout-а. Као што је приказано на слици 3.2, постоји неколико врста електричних веза за међусобно повезивање. Дуге везе (*long*) протежу се целом дужином и ширином FPGA компоненте. *Hex* везе повезују сваки трећи или шести логички блок, у свим правцима. *Double* веза протеже се над највише два логичка блока, у свим правцима док *single* веза повезује само суседне логичке блокове. Сигнали такта повезани су специјалном мрежом за дистрибуцију такта.



Слика 3.2: Типови електричних веза за међусобно повезивање и начини међусобног повезивања логичких блокова

3.2. Мерна поставка

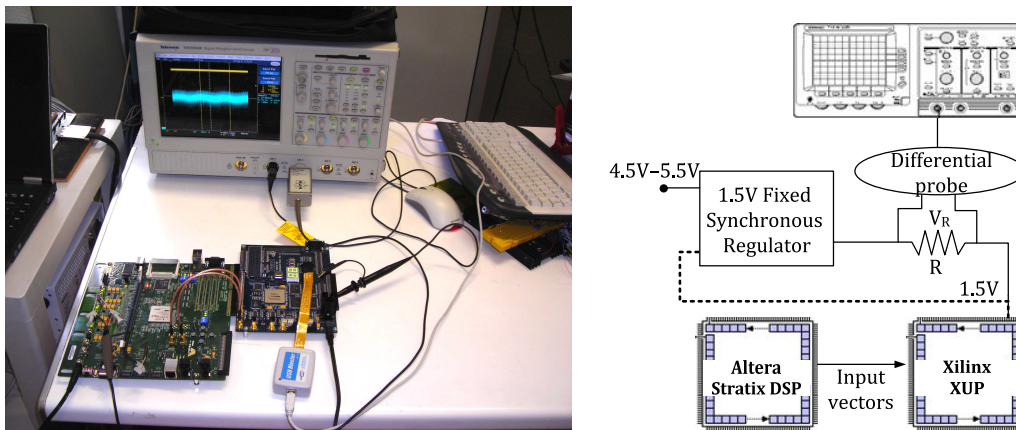
Први циљ методологије мерења потрошње је добијање вредности капацитивности веза коришћених за глобално повезивање у Xilinx Virtex II Pro компонентама будући да ове вредности зависе од технолошких параметара FPGA чипа и да нам нису на располагању. Једном одређене, вредности ових капацитивности користиће се надаље приликом одређивања измерене потрошње интерконекција.

Метода за одређивање капацитивности веза описана је у [Jev09] а базира се на стандардној процедури екстраковања ефективне капацитивности описаној у [EJH+04, DT05].

Мерна поставка заједно са блок шемом представљена је на слици 3.3. Систем се састоји из две FPGA плоче: Xilinx-ове XUP плоче и Stratix DSP Altera-ине развојне плоче. Altera-ина развојна плоча се користи за генерисање симулационих улазних сигнала и за њихово достављање XUP плочи која служи за мерење потрошње одређеног дизајна. Како су на XUP плочи раздвојена напајања језгра, улазно-излазних пинова и помоћно напајање, директно је мерена потрошња језгра FPGA. Извор напајања за језгро од 1.5V добијен је на од стране синхроног *buck-switching* регулатора који се екстерно напаја са 4.5V–5.5V [Xil05].

На улазу извора за најапање језгра у FPGA чип коришћена је отпорност (R), и за сваки тестни дизајн мерен је пад напона на њој што је омогућавало прорачун струје коју „вуче“ FPGA компонента. Вредност отпорности постављена је тако да обезбеди максималну прецизност мерења и да осигура исправну функционалност *buck-switching* PWM (Pulse-Width Modulation) регулатора на XUP плочи. Толеранција отпорности је $\pm 5\%$ што представља систематску грешку мерења. У Додатку А заинтересовани читалац може пронаћи додатне информације о мерном

систему и о томе како се одређује вредност отпорности R . Функционалност самог чипа осигурана је директном повратном спрегом од извора напајања језгра чипа до улаза у регулатор (испрекидана линија на блок шеми са слике 3.3). Вредност напона на отпорнику мерена је уз помоћ диференцијалне сонде Tektronix P6248 осцилоскопа. Да би измерена вредност напона на отпорнику била што тачнија напон је мерен за оба могућа положаја крајева диференцијалне сонде. За сваки положај сонде добија се по 1000000 вредности напона на отпорнику (100 вредности за сваки од 10000 парова улазних вектора). Све измерене вредности напона памте се у *.txt* фајл који се након тога обрађује у MATLAB програмском пакету. Резултат те обраде је просечна вредност напона на отпорнику R . На Altera-иној плочи генерише се и један додатни сигнал који има за циљ да означи почетак и крај секвенце улазних вектора. Вредност потрошње добија се множењем напона напајања језгра (1.5V) и струје која протиче кроз отпорник, а која се добија дељењем добијене вредности напона на отпорнику његовом вредношћу (R).



Слика 3.3: Мерна поставка

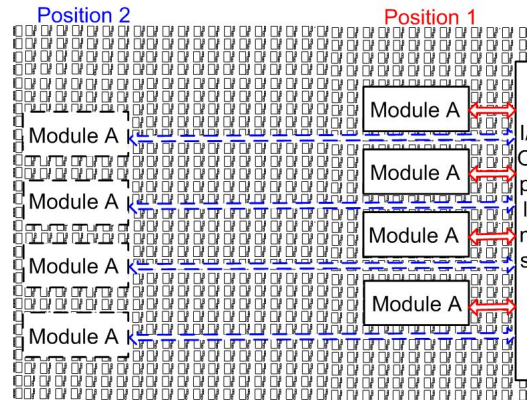
Да би се одредила капацитивност електричних веза мерена је потрошња простих дизајна, углавном IP језгара сабирача или множача (у даљем тексту називаће се модулима) реплицираних неколико пута (од један до четири) како би се побољшала тачност мерења.

Пошто се дизајн стимулише екстерно, не садржи додатне блокове потребне за генерисање улазних сигнала (меморијски низови, контролна логика итд.) који би могли да утичу на измерену укупну потрошњу и да онемогуће раздвајање потрошње модула од потрошњи интерконекција. Према томе, сваки дизајн се састоји од неколико идентичних модула и линија које повезују пинове модула са улазно-излазним (I/O) пиновима (види слику 3.4).

За сваки дизајн понављају се следећи кораци:

1. Најпре се мери статичка потрошња. У том случају дизајн се не побуђује улазним сигнаlima, а нема ни сигнала такта.
2. Након тога се мери потрошња дистрибуционе мреже такта заједно са статичком потрошњом. Улазни синали су идаље сви на нивоу логичке '0'.
3. На крају се мери укупна потрошња дизајна. У том случају улазни вектори имају по 10000 вредности, са Gauss-овом су расподелом и различитим коефицијентима аутокорељације.

Статичка потрошња и потрошња такта одузимају се од укупне потрошње дизајна добијене у кораку 3. За резултат се добија изолована динамичка потрошња логике и глобалних интерконекција.

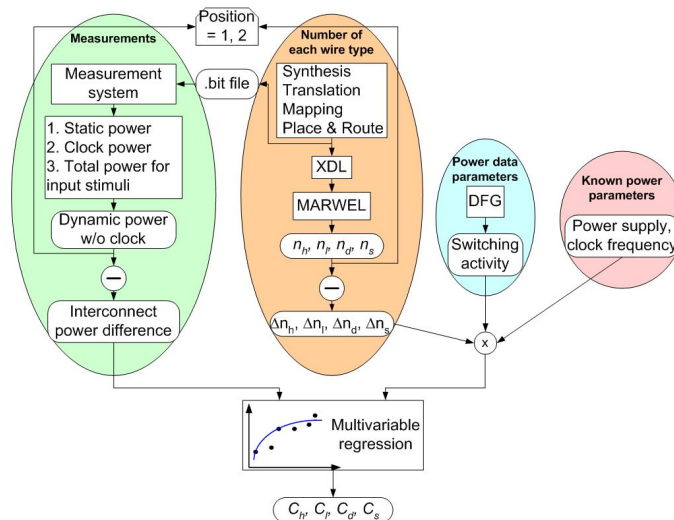


Слика 3.4: Два различита положаја модула унутар FPGA компоненте

Да би се раздвојила потрошња логике од потрошње глобалних интерконекција у [Jev09] коришћена је методологија слична оној у [DT05] где се ефективне вредности капацитивности сваке везе добијају мерењима. Методологија је представљена на слици 3.5.

Најпре се понавља низ мерења описан у корацима 1 до 4, и то за два различита положаја модула на чипу: један када су модули лоцирани близу I/O пинова, и други када су далеко од њих (види слику 3.4). На положај модула у Xilinx ISE софтверу утиче се уз помоћ фајла за корисничка ограничења (user constraint file, .ucf). Одузимањем две вредности за динамичку потрошњу (добијене за два различита положаја модула) добија се вредност потрошње која одговара разлици интерконекција између два дизајна.

Битно је напоменути и то да оба дизајна имају регистроване улазе и излазе. На овај начин елиминишу се гличеве који би се могли десити унутар дизајна услед различитих времена успостављања логичких вредности на улазима.



Слика 3.5: Методологија за екстракцију ефективне капацитивности интерконекција

3.2.1. Екстракција капацитивности електричних веза

Што се тиче потрошње на интерконекцијама, највећи потрошачи су *long* везе, затим *hex* па *double*, док су *single* везе најмањи потрошачи [SKB02, PNB06].

У [Jev09] добијена ефективна вредност за капацитивност сваке електричне везе (*long*, *hex*, *double* и *single*) представља капацитивност везе за повезивање заједно са програмабилним прекидачем који је побуђује. Након размештања и повезивања дизајна Xilinx-ов ISE алат генерише *.ncd* (native circuit description) фајл који представља физички опис дизајна који се имплементира на одређену FPGA компоненту. На Политехничком Универзитету у Мадриду развијен је C++ алат назван MARWEL (Measurement of ARchitectural WirE Length) који је у стању да из XDL (Xilinx Design Language) фајла екстракује информацију о дужини и броју различитих врста веза. XDL фајл је уствари текстуална верзија физичког описа дизајна након размештања и повезивања (*.ncd*), а креира се уз помоћ XDL алата. Више детаља о MARWEL алату читалац може пронаћи у додатку А.

Према томе, за сваку интерконекцију i која иде од или до I/O пинова у дизајну MARWEL даје број *long*, *hex*, *double* и *single* веза које су искоришћене за повезивање: n_{li} , n_{hi} , n_{di} и n_{si} .

Према једначини 1.4, динамичка потрошња интерконекција зависи од четири параметра. Од тих параметара два су већ позната, будући да је вредност напона напајања за Virtex II Pro компоненте једнака 1.5V, а фреквенција такта једнака вредности која се користила приликом мерења (50MHz или 100MHz).

Измерене динамичке потрошње дизајна за две различите позиције модула се међусобно одузимају а као резултат се добија измерена потрошња интерконекција будући да су динамичке потрошње модула једнаке и да се приликом овог одузимања елиминишу.

Према томе, разлика у потрошњи дизајна за две различите позиције може да се изрази као:

$$\begin{aligned}
 P_1 - P_2 = & 0.5 \cdot V_{dd}^2 \cdot f \cdot \left(C_l \cdot \sum_{i=1}^{I_1+I_2+O} [(n_{li}^1 - n_{li}^2) * sw_i] \right. \\
 & + C_h \cdot \sum_{i=1}^{I_1+I_2+O} [(n_{hi}^1 - n_{hi}^2) * sw_i] \\
 & C_d \cdot \sum_{i=1}^{I_1+I_2+O} [(n_{di}^1 - n_{di}^2) * sw_i] \\
 & \left. C_s \cdot \sum_{i=1}^{I_1+I_2+O} [(n_{si}^1 - n_{si}^2) * sw_i] \right)
 \end{aligned} \tag{3.1}$$

где су P_1 и P_2 измерене динамичке потрошње дизајна на позицијама близу и далеко од I/O пинова, респективно; C_l , C_h , C_d , C_s су променљиве које представљају ефективне капацитивности *long*, *hex*, *double* и *single* веза, респективно; I_1 и I_2 су дужине (број битова) два улазна операнда, док O представља број битова излаза. Позиција дизајна дефинисана је индексима 1 (близу) и 2 (далеко). Да би се добиле што тачније вредности за ефективне капацитивности веза извршен је велики број мерења са различитим дужинама улазних и излазних сигнала. У табели 3.1

приказане су ефективне вредности капацитивности сваке везе добијене у [Jev09]. Ове вредности капацитивности биће надаље коришћене и у нашим анализама.

Табела 3.1: Вредности ефективних капацитивности различитих типова веза

Тип везе	Капацитивност по јединици дужине [fF]
Long	178.133
Hex	86.578
Double	71.47
Single	≈0

Када су нам једном познате вредности за ефективне капацитивности веза, имајући информацију о транзиционим активностима и броју веза сваког типа, лако можемо израчунати потрошњу интерконекције:

$$P = 0.5 \cdot V_{dd}^2 \cdot f \cdot sw \cdot (n_l \cdot C_l + n_h \cdot C_h + n_d \cdot C_d + n_s \cdot C_s) \quad (3.2)$$

3.2.2. Потрошња логике и улазних бафера

Кад се од укупне динамичке потрошње дизајна одузме потрошња интерконекција добија се потрошња модула (логике) заједно са потрошњом улазних бафера (будући да се улазни бафери такође напајају напоном напајања језгра). Да бисмо били у стању да добијемо потрошњу модула неопходно нам је и да познајемо ефективну капацитивност улазних бафера. Ова капацитивност је прорачуната мерењем потрошње два дизајна: једног који садржи три множача имплементирана у LUT табеле, и другог који се састоји од само једног LUT множача. Када од измерене потрошње оба одузмемо одговарајућу потрошњу интерконекција добијамо следеће вредности за потрошњу логике:

$$P_{log,1} = 3 * P_{mult} + P_{in_buf} \quad (3.3)$$

$$P_{log,2} = P_{mult} + P_{in_buf}$$

где су $P_{log,1}$ и $P_{log,2}$ потрошње логике првог и другог дизајна, респективно, P_{mult} потрошња логике множача а P_{in_buf} потрошња улазних бафера. Из ове две једначине у стању смо да добијемо потрошњу улазног бафера а самим тим и вредност његове ефективне капацитивности:

$$C_{in_buf} = \frac{P_{in_buf}}{0.5 \cdot V_{dd}^2 \cdot f \cdot SW_{in_buf}} \quad (3.4)$$

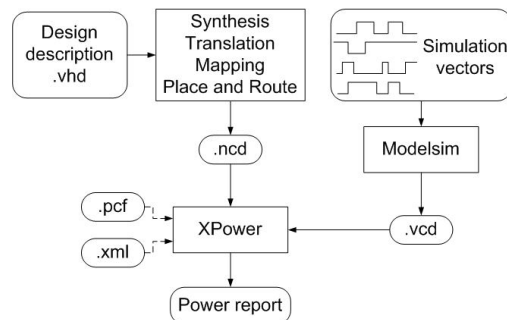
$$SW_{in_buf} = \sum_{i=1}^{N_{in_buf}} sw_in_i$$

где је N_{in_buf} укупан број улаза, а sw_in_i транзициона активност i -тог улаза. Оваквим методом, за ефективну вредност капацитивности улазног бафера добијено је 3.52pF [Jev09].

Познавањем ефективних капацитивности електричних веза и улазног бафера лако можемо од укупне измерене вредности динамичке потрошње одузети потрошњу интерконекција и улазних бафера и добити динамичку потрошњу логике.

3.2.3. XPower алат

XPower Analyzer је Xilinx-ов алат ниског нивоа апстракције дизајна намењен процени потрошње. Типичан ток процене потрошње овим алатом приказан је на слици 3.6. Најпре се врши временска симулација (на нивоу гејтова) дизајна који је имплементиран (размештен и повезан) на FPGA. Резултат симулације је *.vcd* (Value Change Dump, у новијим верзијама ISE алата *.saif* Switching Activity Interchange Format) фајл. За симулације дизајна чија се потрошња процењује коришћен је ModelSim симулатор од Mentor Graphics-a, као један из листе симулатора подржаних од стране XPower-a. *.vcd* фајл садржи детаљне информације о транзиционим активностима свих чворова у дизајну а користи се и као улазни симулациони фајл XPower алата. Поред овог фајла XPower алат користи и *.ncd* фајл који представља физички опис дизајна након размештања и повезивања. Опционо, корисник може да у XPower учита и *.pcf* (Physical Constraints File) фајл који садржи нека ограничења које је дефинисао корисник, као и *.xml* фајл који садржи снимљена подешавања неких параметар алата из претходних XPower анализа.



Слика 3.6: Ток процене потрошње помоћу XPower алата

Излазни фајл из XPower алата је извештај о потрошњи електричне енергије. У зависности од тога које су му информације потребне, корисник може да специфицира да ли ће извештај о потрошњи бити стандардан („Standard“), детаљан („Detailed“) или напредан („Advanced“). Најдетаљније анализе даје напредан извештај, и он је увек специфициран приликом нашег рада са XPower алатом. Извештај о потрошњи садржи потрошњу сваког индивидуалног елемента у дизајну. Потрошње индивидуалних елемената можемо поделити у четири групе:

1. Потрошња дистрибуционе мреже такта која се састоји од тактних веза и тактних бафера (Clock power group);
2. Потрошња логике тј. потрошња унутар конфигурабилних логичких блокова (CLBs) и уграђених блокова (Logic power group);
3. Потрошња сигнала тј. потрошња локалних конекција унутар компоненте (конекције између CLB-ова) и глобалних конекција између I/O пинова и улазних и излазних регистара компоненте (Signals power group);
4. Потрошња улазних бафера (Inputs power group).

Процене потрошње добијене помоћу XPower алата углавном смо користили као референцу. Поредили смо их са измереним вредностима потрошње на плочи

као и са вредностима потрошње које смо добили помоћу нашег модела за процену.

Глава 4

Модели за процену динамичке потрошње логичких структура

ДИЗАЈНИРАЊЕ интегрисаних кола ниске потрошње не би било комплетно и оптимално без алата који су у стању да тачно предвиде и оптимизују потрошњу. Процену потрошње пожељно је обавити што је могуће брже и у свим фазама дизајнирања система. Нарочито су алати за брзу процену потрошње корисни у оквиру оптимизационе петље која је обично итеративна и која има за циљ да потрошњу дизајна „уклопи“ у унапред задати енергетски буџет. Тиме се избегава процес редизајнирања система који је обично скуп и временски захтеван.

У овој глави описан је аналитички приступ процени динамичке потрошње аритметичких кола која се имплементирају на FPGA. Модели за процену потрошње су високог нивоа (информације о дизајну чија се процена потрошње врши су на високом нивоу апстракције) и представљају наставак рада описаног у [Jev09]. Најпре је дат преглед досадашњих активности на пољу процене потрошње (поглавље 4.1). Поступак моделовања улазних сигнала описан је у поглављу 4.2. Показано је како се на основу параметара сигнала на нивоу речи добија транзициона активност сваког бита улазног сигнала. У поглављу 4.3 детаљно је описана методологија одређивања транзиционе активности сваког чвора интерне структуре аритметичког кола. Ревизија модела улазног сигнала за аритметичке компоненте за које се испоставило да испољавају неку врсту нелинеарности на својим излазима (бинари множачи) налази се у поглављу 4.4. На крају су у поглављу 4.5 представљени модели за процену динамичке потрошње различитих типова бинарних делитеља. Такође су представљени и резултати евалуације ових модела.

За све компоненте којима се врши процена потрошње сматра се да имају регистроване улазе и излазе. Предложена методологија за процену потрошње базира се на аналитичком моделу транзиционе активности компоненте. Модел за процену потрошње је параметризован, а параметри од којих зависи су дужина

(број битова) и статистика (средња вредност, варијација, аутокорељација) сигнала операнада. Треба рећи и то да се модели описани у дисертацији односе на имплементације дизајна у LUT табелама унутар FPGA компоненте. Процена потрошње дизајна који се имплементирају у уграђеним (embedded) блоковима незнатно се разликује зато што је структура ових блокова ближа ASIC структури будући да су уграђени блокови специјализовани за обављање одређене аритметичке операције и да нема програмабилних интерконекија.

Према једначини 1.4 динамичка потрошња зависи од четири параметра: V_{dd} , f , SW и C_l . Величина напона напајања језгра и радна фреквенција дизајна су константни параметри. Будући да су DSP блокови углавном изграђени од основних логичких елемената (полусабирача, сабирача кола за множење (AND) итд.), а такође и због чињенице да FPGA компоненте имају регуларну структуру (изграђене су од логичких блокова са LUT табелама) можемо претпоставити да је ефективна капацитивност FPGA логичких блокова, C_l , константна. Да бисмо могли да проценимо динамичку потрошњу дизајна неопходно нам је да израчунамо једини непознати параметар, суму транзиционих активности дизајна – SW :

$$P_D = C_l \cdot V_{dd}^2 \cdot f \cdot \sum \alpha = a \cdot SW \quad (4.1)$$

где је константа a параметар који се одређује емпиријски на начин који ће касније бити објашњен.

У наставку је, након прегледа техника за процену потрошње, представљен модел улазних сигнала који има за задатак да са што је могуће мање параметара моделује утицај улазних сигнала (тј. сигнала операнада) на динамичку потрошњу дизајна. Након тога је описан начин рачунања транзиционе активности који користи претходно описани модел улазних сигнала а базиран је на ширењу транзиционе активности кроз логичку структуру дизајна.

4.1. Преглед техника за процену потрошње

Први EDA (Electronic Design Automation) софтверски пакети за аутоматизацију процеса пројектовања интегрисаних кола били су опремљени различитим алаткама намењеним за симулацију (предвиђање понашања) и анализу њихових перформанси (брзина, заузеће ресурса чипа, детектовање дефеката итд.). Кад је потрошња електричне енергије постала битан параметар интегрисаног кола, многи EDA пакети својој палети алатки додали су и алате за њену процену.

Модели за процену потрошње међусобно се разликују по природи електричне енергије коју процењују (статичка, динамичка, потрошња кратког споја или укупна потрошња), као и по нивоу апстракције дизајна чија се потрошња процењује. Што је ниво апстракције дизајна већи, процене потрошње су брже али и мање тачне. Постоје два приступа проблему процене електричне енергије: статистички и пробабилистички [Mac08].

Статистички методи током симулације дизајна вектором улазних података прикупљају статистичке податке о сваком чвору у дизајну. Најпростије статистичке технике за процену потрошње представљене су у [Den94, Lan94, Sch95, Geo94]. Ове технике су тачне али процесорски и временски захтевне (нарочито за комплексна кола) и зависне од вектора улазних података потребних за симулацију. За решавање овог последњег проблема, статистички приступи

базирани на Monte Carlo симулацији представљени су у [BNYT93, Tod02, TB05]. Под претпоставком да потрошња кола током дугог временског интервала T има нормалну расподелу, на примарне улазе кола примењују се случајно генерисани улазни вектори и посматра се потрошња кола током временског интервала T .

Са друге стране, пробабилистички методи анализирају коло и генеришу аналитичке пробабилистичке изразе за сваки чвор у колу, почев од примарних улаза, па све до излаза кола [Mac08, Cir87, CRP94, MRT08]. Према томе, код пробабилистичких метода процена потрошње не зависи од димензије вектора улазних сигнала потребних за симулацију већ само од њихове статистике. И ови методи, слично статистичким, имају проблеме приликом анализе комплексних кола будући да се са повећањем броја улаза у коло и његове логичке дубине повећава и комплексност аналитичких пробабилистичких израза.

Неки неконвенционални приступи у процени потрошње могу се наћи у [Mak00]. За процену на ниском нивоу апстракције дизајна аутор приликом одређивања укупне транзиционе капацитивности кола користи сопствени програмски језик (AleC++) и симулатор (AleCsis). За сваки логички елемент унутар дизајна потребно је поседовати модел на ниском нивоу апстракције са информацијама о капацитивностима елемента. Будући да је на ниском нивоу апстракције дизајна, модел је спор приликом анализе комплексних кола. Још један приступ, такође представљен у [Mak00], базиран је на интегралнеу таласног облика струје нападања кола. Будући да се оваквим приступом најтачније може проценити укупна потрошња кола, аутор за одређивање површине струјног импулса користи неуронску мрежу са 3 слоја.

Опис техника за процену статичке и потрошње кратког споја заинтересовани читалац може пронаћи у [HDC11, NS00], респективно.

Што се тиче естимације дизајна имплементираних на FPGA, постојећи модели за процену потрошње теже да потрошњу дизајна представе у форми једначине. Променљиве у једначини зависе од различитих фактора (статистике улазних и излазних сигнала, дужине операнада, fanout-а кола, структуре компоненти итд.). Неки приступи у процени потрошње FPGA дизајна представљени су у [CCF03, CJMP03, ESJ06, AFJS07, AN04]. Грешке у процени потрошње крећу се у опсегу од 10% до 30%. Међутим, док неки од метода нису поређени са измереним вредностима потрошње [CCF03], други су екстремно временски захтевни (чак и до 12 сати) [CJMP03, AN04] или захтевају дуге процедуре калибрације [ESJ06, AFJS07].

На крају, постији неколико алата дизајнираних за комерцијалне FPGA компоненте. Најпознатији су Xilinx-ов *XPower* и Altera-ин *PowerPlay*. Ови алати обезбеђују детаљан извештај о потрошњи дизајна који је базиран на капацитивностима и броју искоришћених логичких блокова као и транзиционој активности у свакој тачки дизајна. Почетне верзије ових алата имале су ограничену тачност приликом поређења са измереним вредностима потрошње [ESJ06]. Касније верзије су се све више и више усавршавале. Међутим, чак се и ови комерцијално доступни алати суочавају са проблемима (постају процесорски и временски захтевни) приликом анализе комплексних дизајна са много сигнала.

4.2. Моделовање улазног сигнала

Потрошња дизајна директна је функција броја транзиција сигнала које се дешавају у његовим капацитивним чворовима. Просечан број транзиција сигнала

(промена логичког стања) у неком чвору интегрисаног кола за време једног тактног интервала изражава се преко транзиционе активности и вероватноће појаве транзиције. Колика ће бити транзициона активност неког чвора у дизајну зависи од сигнала који се доводе на улазне прикључке дизајна. Према томе, потрошња дизајна који се имплементира на FPGA значајно може да варира са променом статистичке расподеле улазних података.

Транзициона активност зависи од текуће и прве претходне вредности улазних сигнала. Најтачнији (али и временски најзахтевнији) метод за рачунање транзиционе активности је симулација имплементираних (размештеног и повезаног) дизајна реалним улазним подацима. Ипак, будући да је циљ дисертације моделовање потрошње на високом нивоу апстракције дизајна, неопходно је извршити моделовање улазних сигнала малим бројем параметара али тако да модел буде у стању да изрази утицај улазних података на динамичку потрошњу дизајна. Уместо анализе транзиционе активности на нивоу бита, многе технике процене потрошње, да би израчунале транзициону активност у колу, користе статистику улазног сигнала на нивоу речи. Овакав приступ коришћен је и у овој дисертацији. Код модела сигнала на нивоу речи улазни сигнали представљени су својим статистичким параметрима: средњом вредношћу, варијансом, коефицијентом аутокорељације итд.

Будући да разматрамо DSP компоненте (кола за процесирање дигиталног сигнала), претпоставка за улазне сигнале је да су стационарни и да имају Gauss-ову (нормалну) расподелу. Ова претпоставка може се сматрати исправном будући да је Gauss-ова расподела често коришћена за тачно моделовање DSP сигнала [LR95, JKS+99, CGC05].

4.2.1. Сигнали чија је средња вредност једнака 0

Најпре ће бити описан модел за генерисање Гаусовог (Gauss) сигнала [CGC05] будући да ће се он надаље користити за рачунање неких битних параметара модела улазног сигнала, а такође и за испитивање утицаја променљиве статистике улазних сигнала на потрошњу дизајна. Модел за генерисање Гаусовог сигнала у стању је да истовремено генерише два сигнала са Гаусовом расподелом и изабраним статистикама.

Транзициона активност у синхроним колима у потпуности је дефинисана тренутном и првом претходном вредношћу сигнала. Према томе, за аритметичка кола са два операнда x и y , под претпоставком да су сигнали операнда стационарни, са средњом вредношћу једнаком 0 и Гаусовом расподелом, транзициона активност зависиће од њихове заједничке функције густине вероватноће (PDF – probability density function) $p(x_0, x_1, y_0, y_1)$. Ова функција је за сигнале са средњом вредношћу једнаком нули и Гаусовом расподелом једнака [Pap91]:

$$p(x) = \frac{1}{(2\pi)^{n/2} \cdot \det^{1/2}(C)} \cdot \exp\left[-\frac{1}{2} \cdot x^T \cdot C^{-1} \cdot x\right] \quad (4.2)$$

где је $x = [x_1, x_2, \dots, x_n]$ вектор сигнала а C симетрична матрица димензија $n \times n$ са $C_{i,j} = E\{x_i x_j\}$ ($E\{\}$ је функција математичког очекивања). За два улазна сигнала добијамо следећи израз за њихову PDF функцију:

$$p(x_0, x_1, y_0, y_1) = \frac{1}{(4\pi^2) \cdot \det^{1/2}(C)} \cdot \exp\left[-\frac{1}{2} \cdot [x_0 x_1 y_0 y_1] \cdot C^{-1} \cdot [x_0 x_1 y_0 y_1]^T\right] \quad (4.3)$$

где је матрица C једнака:

$$C = \begin{bmatrix} r_{xx0} & r_{xx1} & r_{xy0} & r_{xy1} \\ r_{xx1} & r_{xx0} & r_{yx1} & r_{yx0} \\ r_{xy0} & r_{yx1} & r_{yy0} & r_{yy1} \\ r_{xy1} & r_{yx0} & r_{yy1} & r_{yy0} \end{bmatrix}. \quad (4.4)$$

Према томе, све информације потребне за карактеризацију транзиционе активности садржане су у следећих седам статистичких параметара:

- 1) r_{xx0} , варијанса сигнала x
- 2) r_{xx1} , коефицијент аутокорељације сигнала x
- 3) r_{xy0} , међусобна корелација два сигнала
- 4) r_{xy1} , међусобна корелација два сигнала са *unit-time lag* у y
- 5) r_{yx1} , међусобна корелација два сигнала са *unit-time lag* у x
- 6) r_{yy0} , варијанса сигнала y
- 7) r_{yy1} , коефицијент аутокорељације сигнала y

Два сигнала (x и y) са Гаусовом расподелом и седам изабраних статистичких параметара могу се генерисати из два временски и просторно некорелисана Гаусова сигнала (u и v) са средњом вредношћу једнаком нули на следећи начин:

$$p = \beta \cdot u \quad (4.5)$$

$$q = \delta \cdot u + \gamma \cdot v \quad (4.6)$$

Након тога, елементи x и y низа добијају се као:

$$x_i = p_i + a_1 \cdot x_{i-1} + a_3 \cdot y_{i-1} \quad (4.7)$$

$$y_i = q_i + a_2 \cdot x_{i-1} + a_4 \cdot y_{i-1} \quad (4.8)$$

где су $a_1, a_2, a_3, a_4, \gamma, \delta$ и β коефицијенти који се добијају из следећих израза:

$$\begin{bmatrix} r_{xx1} \\ r_{xy1} \end{bmatrix} = \begin{bmatrix} r_{xx0} & r_{xy0} \\ r_{xy0} & r_{yy0} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_3 \end{bmatrix} \quad (4.9)$$

$$\begin{bmatrix} r_{yx1} \\ r_{yy1} \end{bmatrix} = \begin{bmatrix} r_{xx0} & r_{xy0} \\ r_{xy0} & r_{yy0} \end{bmatrix} \cdot \begin{bmatrix} a_2 \\ a_4 \end{bmatrix} \quad (4.10)$$

$$r_{xx0} = r_{pp0} + a_1 \cdot r_{xx1} + a_3 \cdot r_{xy1} \quad (4.11)$$

$$r_{yy0} = r_{qq0} + a_2 \cdot r_{yx1} + a_4 \cdot r_{yy1} \quad (4.12)$$

$$r_{xy0} = r_{pq0} + a_1 \cdot r_{yx1} + a_3 \cdot r_{yy1} \quad (4.13)$$

$$r_{pp0} = \beta^2, r_{qq0} = \gamma^2 + \delta^2, r_{pq0} = \beta \cdot \delta \quad (4.14)$$

У [CGC05] показано је да динамичка потрошња у аритметичким колима много више зависи од коефицијента аутокорелације него од коефицијента међусобне корелације. Према томе, међусобна корелације неће бити један од параметара за моделовање улазног сигнала. Тиме ће се број статистичких параметара потребних за одређивање потрошње свести на: r_{xx0} , r_{yy0} , r_{xx1} , r_{yy1} .

Коришћењем једначине 4.7, уз претпоставку да коефицијенти међусобне корелације могу бити занемарени, може се видети да се сигнал са Гаусовом расподелом и неким одређеним коефицијентом аутокорелације може представити на следећи начин:

$$x(n) = \beta \cdot u(n) + a_1 \cdot x(n-1) \quad (4.15)$$

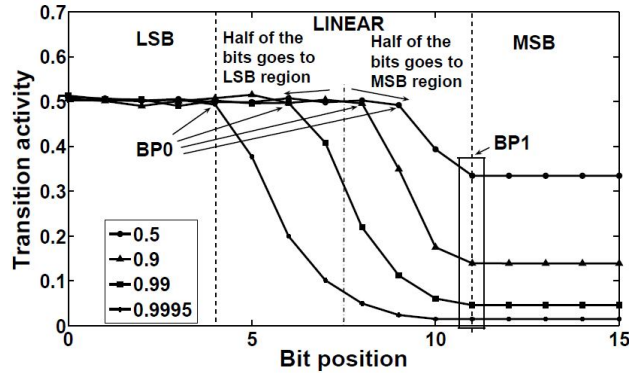
Ова једначина биће коришћена приликом одређивања неких значајних карактеристика модела сигнала.

Модел сигнала на нивоу речи

У [LR95] први пут је примећено да сигнал са Гаусовом расподелом и средњом вредношћу једнаком нули може бити подељен на три региона, у зависности од статистике сигнала на нивоу речи. На слици 4.1, за различите вредности коефицијената аутокорелације, приказана је транзициона активност битова сигнала у зависности од њихове позиције у речи. Сви сигнали имају Гаусову расподелу, средњу вредност једнаку нули и исту варијансу σ^2 . Три поменута региона су: 1) регион LSB некорелисаних битова, 2) линеарни регион са корелисаним битовима и 3) MSB (Most Significant Bits) битови знака. Некорелисани битови сигнала имају фиксну транзициону активност од 0.5 и почињу од бита најмање тежине (Least Significant Bit - LSB) па све до одређене границе ($BP0$). Јако корелисани MSB битови простиру се од бита највеће тежине па све до друге границе ($BP1$). Транзиционе активности битова у средњем региону линеарно повезују MSB битове и некорелисане битове. Једначине које дефинишу границе $BP0$ и $BP1$ и зависе од статистичких параметара сигнала биће дате у наставку.

Улазни сигнали аритметичких кола чија се потрошња процењује моделовани су управо описаном DBT (dual-bit type) методом која је у [RSH97, SP00] проширена и побољшана тачнијим изразима за границе $BP0$ и $BP1$ (будући да су у њиховим анализама сигнали представљени ARMA (autoregressive moving-average) моделима. Такође, уместо мерења транзиционе густине MSB битова (што је рађено у [LR95]), она је у [RSH97, SP00] процењена на основу статистике сигнала на нивоу речи.

Дакле, уместо да се подела на регионе у оквиру речи врши на основу транзиционе активности, као што је рађено у [LR95], у дисертацији ће, следећи методологију описану у [RSH97], подела на регионе бити извршена на основу корелације на нивоу бита, ρ_i .



Слика 4.1: Транзициона активност бита у зависности од њихове позиције у речи за сигнале са средњом вредношћу једнаком 0

По дефиницији, $\rho_i=0$ за $i < BP0$. За MSB битеве знака може се увести следећа претпоставка: $\rho_{BP1}=\rho$, где је ρ временска корелација на нивоу речи.

Изрази за коефицијенте аутокорелације сва три региона су следећи:

$$\rho_i = \begin{cases} 0 & i < BP0 \\ \frac{(i - BP0 + 1) \cdot \rho_{BP1}}{BP1 - BP0} & BP0 \leq i < BP1 \\ \rho_{BP1} & i \geq BP1 \end{cases} \quad (4.16)$$

За прорачунавање тачне вредности транзиционе активности, sw_i , користи се следећа релација између вероватноће на нивоу бита, p_i , и корелације на нивоу бита, ρ_i :

$$sw_i = 2 \cdot p_i \cdot (1 - p_i) \cdot (1 - \rho_i) \quad (4.17)$$

Вероватноћа на нивоу бита, p_i , (вероватноћа да је бит у одређеном логичком стању) једнака је:

$$p_i = \sum_{j \in \Psi} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-(j-\mu)^2 / 2\sigma^2} \quad (4.18)$$

где је Ψ скуп свих вредности које сигнал може имати под условом да му је i -ти бит једнак 1.

Будући да вредност сигнала x лежи у интервалу $x_{max}=x+3\sigma$ и $x_{min}=x-3\sigma$, динамички опсег сигнала једнак је $x_{max}-x_{min}$, и за његово представљање потребно је $\log_2(x_{max}-x_{min})$ бита. Према томе, коначан израз за границу $BP1$ је:

$$BP1 = \lceil \log_2(6\sigma) \rceil \quad (4.19)$$

где је $\lceil \cdot \rceil$ оператор целобројног заокруживања. Важно је уочити да у изразу за коефицијент аутокорелације на нивоу бита (4.16) MSB регион почиње од бита на позицији $BP1$, што значи да и тај бит такође припада региону знака. Према томе, варијација сигнала представља се са следећим бројем бита:

$$BP1 - 1 = \lceil \log_2(3\sigma) \rceil \quad (4.20)$$

Израз за границу $BP0$ зависи од коефицијената ARMA модела сигнала и за сигнале са Гаусовом расподелом и средњом вредношћу једнаком 0 еквивалентан је:

$$BP0 = \left\lceil \log_2(\sqrt{1 - \rho^2} \cdot \sigma_x) \right\rceil \quad (4.21)$$

Детаље око извођења овог израза заинтересовани читалац може наћи у [RSH97].

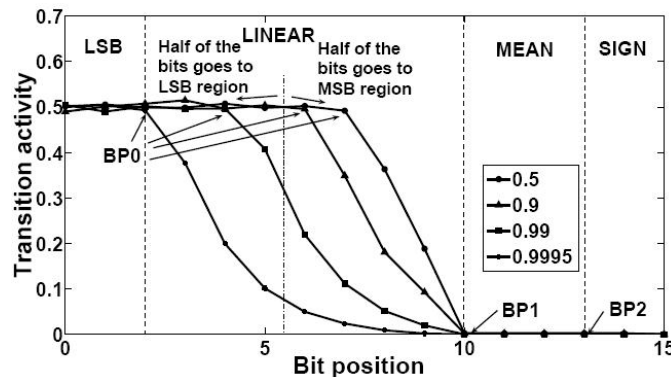
На крају, ради упрошћавања модела, линеарни регион подељен је на два једнака дела. Горња половина придружена је MSB региону, док се доња половина битова линеарног региона придружује LSB битовима (види слику 4.1). Ова апроксимација упрошћава даља рачунања а да се при том не уноси нека значајна грешка у процени транзиционе активности. Иста апроксимација коришћена је и у [LR95, CGC+06].

4.2.2. Сигнали средње вредности различите од 0

Улазни сигнали многих DSP логичких блокова из реалних апликација (у даљем тексту DSP сигнали) имају само позитивне вредности (нпр. некомпресовани сигнали слике где се сваки пиксел представља позитивном целобројном вредношћу). Због тога су у разматрање узети и Гаусови сигнали са средњом вредношћу μ , варијансом σ^2 и коефицијентом аутокорељације ρ .

На слици 4.2, за сигнале са различитим коефицијентом аутокорељације и одређеном средњом вредношћу μ , приказана је транзициона активност битова сигнала у зависности од њихове позиције у речи. И у случају кад средња вредност сигнала није једнака нули, транзициона активност сигнала може се поделити на неколико региона.

Две границе $BP0$ и $BP1$ које су делиле реч сигнала са средњом вредношћу једнаком 0 присутне су и у случају кад је средња вредност сигнала различита од нуле. За сигнал са средњом вредношћу једнаком нули ове две границе биле су довољне да би се прорачунао утицај транзиционе активности сваког региона на укупну транзициону активност компоненте. За сигнале са Гаусовом расподелом и средњом вредношћу различитом од нуле у обзир се морају узети још неки ефекти.



Слика 4.2: Транзициона активност битова у зависности од њихове позиције у речи за сигнале са средњом вредношћу различитом од 0

Регион након границе $BP1$ трансформише се у два под-региона са *mean* битовима и битовима знака (*sign*) (види слику 4.2). Иако је транзициона активност ових битова једнака нули, без обзира на коефицијент аутокорељације, вредности ових битова зависе од вредности и знака средње вредности сигнала, и као такви, имају велики утицај на потрошњу. Овај ефекат је нарочито значајан у

случају разматрања потрошње множача код којих се MSB битови (битови знака) једног од улаза знаковно проширују приликом сабирања парцијалних производа два операнда. У случају да је средња вредност сигнала мања од нуле, битови знака биће једнаки логичкој '1' и утицаће на пропагацију транзиционе активности другог операнда кроз структуру множача. Са друге стране, ако је средња вредност сигнала позитивна (MSB битови једнаки '0'), транзициона активност на излазима одговарајућих елемената у структури биће једнака нули, без обзира на транзициону активност другог операнда. Исти закључак важи и за вредности *mean* битова. Према томе, очигледно је да вредности *mean* битова и битова знака, иако не испољавају транзициону активност, значајно утичу на укупну потрошњу компоненте.

Вредности за границе *BP0* и *BP1* добијају се из израза добијених у претходном поглављу. Нова граница *BP2* и добија се на следећи начин. Максимална вредност сигнала чија средња вредност није једнака нули је $\mu + 3\sigma$ па је за њену бинарну репрезентацију потребно $\log_2(\mu + 3\sigma)$ битова. Као што је малочас поменуто, за представљање варијације сигнала око средње вредности потребно је $v = \log_2(3\sigma)$ битова (види једначину 4.20). Према томе, број битова у сигналу који не мењају своју вредност и чије вредности су једнаке са *K* горњих (MSB) битова бинарне репрезентације средње вредности сигнала једнак је:

$$K = m - v = \log_2(\mu + 3\sigma) - \log_2(3\sigma) \quad (4.22)$$

Ако сигнал има *N* битова, број битова у региону знака који су сви једнаки '0' или '1', у зависности од знака средње вредности сигнала, може се изразити као:

$$S = N - m = N - \log_2(\mu + 3\sigma) \quad (4.23)$$

Као последица тога, MSB регион у случају сигнала са средњом вредношћу различитом од нула чини само половина битова из линеарног региона (види слику 4.2). Трећа гранична вредност (*BP2*) која раздваја *mean* регион од региона знака рачуна се према:

$$BP2 = \lceil \log_2(\mu + 3\sigma) \rceil \quad (4.24)$$

Транзиционе активности битова у сваком региону рачунају се према једначини 4.18. Коефицијент аутокорејације на нивоу бита једнак је 0 за LSB битове, може се апроксимирати са ρ за MSB битове и једнак је 1 за *mean* битове и битове знака.

4.3. Прорачун транзиционе активности

Тек кад нам је позната транзициона активност улазних битова у стању смо да израчунамо транзициону активност унутрашњих чворова у структури компоненте. У наставку овог поглавља биће дат кратак преглед техника за прорачун транзиционе активности праћен описом методологије која је коришћена у овој дисертацији.

4.3.1. Технике прорачуна транзиционе активности

Транзициона активност у чворовима дизајна чија се потрошња процењује зависи како од улазних података тако и од логичке функције дизајна и гличева, нежељених активности које се дешавају услед различих времена простирања сигнала унутар дизајна.

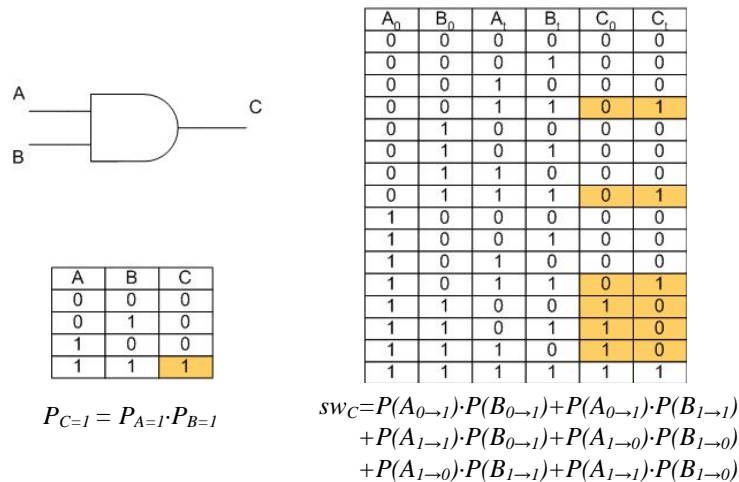
Тачност прорачуна транзиционе активности зависи од временског модела који се користи за моделовање кашњења у колу. Постоје три различита временска модела: модел нултог кашњења, модел променљивог кашњења и модел реалног кашњења.

Модел нултог кашњења претпоставља да је кашњење било ког логичког гејта у дизајну једнако нули тј. да сигнали који се доведу на улаз компоненте моментално произведу излазни сигнал. Овај модел има најмању тачност али и две битне предности. Прво, може се користити на високим нивоима апстракције система (тј. пре размештања и повезивања дизајна), будући да се кашњење на везама не узима у обзир. Друго, ефекат глича није укључен у овај модел. У том случају транзициона активност је у опсегу $[0,1]$ и може се моделовати као вероватноћа да ће сигнал променити своје логичко стање за време једног тактног интервала. У овом случају на прорачун транзиционе активности може се применити пробабилистичка методологија.

Други временски модел је модел променљивог кашњења. Он узима у обзир кашњења сигнала кроз логичке елементе у дизајну, али не и кашњења сигнала на везама (тј. коначно време простирања сигнала). Тачност овог модела је већа а модел се такође може применити на вишим нивоима апстракције. Главни недостаци овог модела су неопходност познавања технолошких и имплементационих детаља дизајна потребних за екстракцију временских параметара, отежана примена пробабилистичке методологије и значајно повећано време прорачуна транзиционе активности.

Трећи временски модел је модел реалног кашњења и узима у обзир сва кашњења у колу. Може се користити само након фаза размештања и повезивања зато што су тек након ових фаза позната кашњења на везама. Иако је највеће тачности, овај модел тешко је применљив на компликованије дизајне будући да време потребно за прорачун постаје екстремно дуго.

Два су различита приступа у решавању проблема процене пропагације транзиционе активности: статистички и пробабилистички. Статистички приступи симулирају дизајн улазним векторима и врше статистичку анализу прикупљајући статистичке податке за сваки чвор у дизајну. Недостатак оваквих приступа је изузетно дуго време анализе, посебно за веће дизајне и дуже симулационе векторе. Са друге стране, пробабилистички методи анализирају коло и генеришу изразе за транзициону активност која се пропагира кроз дизајн. Према томе, пробабилистички методи не зависе од дужине симулационих вектора већ само од њихове статистике (средња вредност, варијанса аутокорељација итд.). И ови методи имају проблема приликом анализе већих кола будући да комплексност аналитичких израза зависи од броја улаза и логичке дубине кола.



Слика 4.3: Пробабилитички метод примењен на логичко I коло

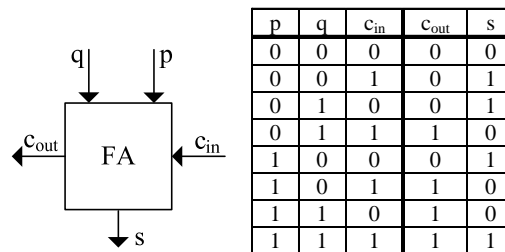
У [TGS+02, TB05] коришћен је статистички приступ за процену потрошње комбинационих кола имплементираних на FPGA. Коришћене су детаљне информације о дизајну које су на располагању након фази размештања и повезивања и применом Monte-Carlo симулација надгледана је активност по тактном интервалу сваког чвора у дизајну и то за насумичне вредности улазних сигнала.

Пробабилитички методи третирају транзициону активност као вероватноћу да ће сигнал променити своје логичко стање за време једног тактног интервала. Поред транзиционе активности, ови методи баратају и са вероватноћама да сигнал има вредност логичке '0' или логичке '1'. Сваки логички гејт представљен је једначином која описује зависност транзиционе активности и вероватноће излазног бита од транзиционих активности и вероватноћа улазних битова гејта. Ова једначина добија се на основу табеле истинитости логичког гејта. На пример, за логичко I коло (види слику 4.3), излаз ће бити у стању логичке '1' само онда кад су оба улаза такође '1'. Према томе, вероватноћа да је излазни бит у стању логичке '1' једнака је производу вероватноћа да су оба улаза у стању '1', у случају да су улази међусобно независни. Прорачун транзиционе активности одвија се слично само што у том случају треба разматрати два сукцесивна временска тренутка. За све комбинације улазних битова, тамо где излаз има различита логичка стања у два сукцесивна временска тренутка дешава се транзиција (тј. промена логичког стања). На слици 4.3 такве комбинације су осенчане. Транзициона активност ће, дакле, зависити како од вероватноћа улазних битова тако и од њихових транзиционих активности. За прорачун транзиционе активности у овој дисертацији биће коришћен пробабилитички метод описан у наставку.

4.3.2. Методологија рачунања транзиционе активности

Имајући у виду да је ћелија потпуног сабирача (види слику 4.4) градивни елемент многих DSP дизајна, на примеру ове ћелије демонстрираћемо методологију прорачуна транзиционе активности. Тачније, фокусираћемо се на прорачун транзиционе активности $carry_{out}$ излаза, као компликованијег за прорачун. Листа појмова коришћених у изразима је следећа:

- p је транзициона вероватноћа (вероватноћа да ће се десети транзиција) једног улаза ћелије потпуног сабирача
- q је транзициона вероватноћа другог улаза ћелије потпуног сабирача
- c_{in} је транзициона вероватноћа улазног *carry* бита
- c_{out} је транзициона вероватноћа излазног *carry* бита
- s је транзициона вероватноћа излаза ћелије потпуног сабирача
- p^0 и p^1 су вероватноће да је улаз p у стању логичке '0' и логичке '1', респективно
- q^0 и q^1 су вероватноће да је улаз q у стању логичке '0' и логичке '1'
- c_{in}^0 и c_{in}^1 су вероватноће да је улазни *carry* бит у стању логичке '0' и логичке '1'
- c_{out}^0 и c_{out}^1 су вероватноће да је излазни *carry* бит у стању логичке '0' и логичке '1'
- s^0 и s^1 су вероватноће да је излаз s у стању логичке '0' и логичке '1'



Слика 4.4: Ћелија потпуног сабирача (Full-Adder, FA) заједно са табелом истинитости

Као што је поменуто раније, транзициона активност зависи од две узастопне вредности сигнала. Излазни битови ћелије потпуног сабирача зависе од три улазна сигнала. Стога, у једном тактном интервалу, постоји максимално 2^3 различитих комбинација тј. стања у којима улази могу да се нађу. За сваку улазну комбинацију, постоји такође 2^3 различитих комбинација за стања улаза у следећем тактном интервалу. На пример, на слици 4.5 је претпостављено да су улази потпуног сабирача у временском тренутку t_0 једнаки „001“. У временском тренутку t_0+1 улази могу бити у једном од осам различитих стања приказаних у табели на слици 4.5. Ипак, само четири комбинације узроковаће транзицију на *carry* излазу c_{out} . Битови који припадају овим комбинацијама и који су у односу на временски тренутако t_0 променили своје логичко стање тамније су осенчани у односу на битове који су задржали логичко стање из временског тренутка t_0 . Свака од ових комбинација дешава се са одређеном вероватноћом. Према томе, у глобалу, требало би разматрати свих $2^3 \times 2^3 = 64$ комбинација улаза у два узастопна временска интервала и међу тим комбинацијама тражити оне које утичу на промену стања излазног бита c_{out} . Да би редуковали број комбинација, разматраћемо све улазне комбинације у временском тренутку t_0 , али ћемо поделити догађаје који могу да се десе у следећем временском тренутку (тактном интервалу) t_0+1 на четири могућа случаја. У првом случају, ниједан од улаза не мења своје логичко стање у следећем тактном интервалу. Према томе, неће бити транзиционе активности на *carry* излазу c_{out} . У другом случају, једино један од улаза мења своје логичко стање. У случају примера са слике 4.5 за комбинације „000“ и „111“ у тренутку t_0+1 неће бити промене логичког стања на излазном *carry* биту. За комбинације „001“, „010“ и „100“, ако се мења један од улаза у стању логичке нуле десиће се транзиција и на c_{out} биту. Према томе, транзициона активност за овај случај је:

Временски тренутак t_0+1

p	q	C_{in}	C_{out}
0	0	1	0

p	q	C_{in}	C_{out}
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

← Други случај

← Други случај

← Четврти случај

← Трећи случај

Слика 4.5: Пробабилистичка методологија за прорачун транзиционе активности

$$\begin{aligned}
 c_{out}^{A1} &= p^0 \cdot q^0 \cdot c_{in}^1 \cdot (1 - c_{in}) \cdot (p + q - 2 \cdot p \cdot q) + \\
 & p^1 \cdot q^0 \cdot c_{in}^0 \cdot (1 - p) \cdot (q + c_{in} - 2 \cdot q \cdot c_{in}) + \\
 & p^0 \cdot q^1 \cdot c_{in}^0 \cdot (1 - q) \cdot (p + c_{in} - 2 \cdot c_{in} \cdot p)
 \end{aligned} \tag{4.25}$$

За остале комбинације из скупа овог другог случаја („011“, „110“ и „101“) промена логичког стања било које јединице са улаза утицаће на транзициону активност *carry* бита:

$$\begin{aligned}
 c_{out}^{A2} &= p^0 \cdot q^1 \cdot c_{in}^1 \cdot (1 - p) \cdot (q + c_{in} - 2 \cdot q \cdot c_{in}) + \\
 & p^1 \cdot q^0 \cdot c_{in}^1 \cdot (1 - q) \cdot (p + c_{in} - 2 \cdot c_{in} \cdot p) + \\
 & p^1 \cdot q^1 \cdot c_{in}^0 \cdot (1 - c_{in}) \cdot (p + q - 2 \cdot q \cdot p)
 \end{aligned} \tag{4.26}$$

У трећем случају, у следећем тактном интервалу мењају се логичка стања два улаза док трећи остаје константан. У овом случају, за комбинације бита „000“ и „111“ промена стања било која два операнда утицаће на транзицију излаза C_{out} . Према томе, транзициона активност за ове комбинације је:

$$c_{out}^{B1} = (p^0 \cdot q^0 \cdot c_{in}^0 + p^1 \cdot q^1 \cdot c_{in}^1) \cdot (p \cdot q \cdot (1 - c_{in}) + (1 - p) \cdot q \cdot c_{in}) \tag{4.27}$$

Комбинације које сарже две нуле и једну јединицу, као „001“, „010“ и „100“ узроковаће транзицију на излазном *carry* биту ако улази који су у стањима логичке нуле промене своје логичко стање. Проузрокована транзициона активност је:

$$\begin{aligned}
 c_{out}^{B2} &= p^0 \cdot q^0 \cdot c_{in}^1 \cdot p \cdot q \cdot (1 - c_{in}) + \\
 & p^0 \cdot q^1 \cdot c_{in}^0 \cdot (1 - q) \cdot p \cdot c_{in} + \\
 & p^1 \cdot q^0 \cdot c_{in}^0 \cdot (1 - p) \cdot q \cdot c_{in}
 \end{aligned} \tag{4.28}$$

На крају, комбинације „011“, „110“ и „101“ узроковаће транзицију на C_{out} ако се обе јединице са улаза промене у логичке нуле. Према томе, имамо:

$$\begin{aligned}
 c_{out}^{B3} &= p^0 \cdot q^1 \cdot c_{in}^1 \cdot (1 - p) \cdot q \cdot c_{in} + \\
 & p^1 \cdot q^0 \cdot c_{in}^1 \cdot p \cdot (1 - q) \cdot c_{in} + \\
 & p^1 \cdot q^1 \cdot c_{in}^0 \cdot p \cdot q \cdot (1 - c_{in})
 \end{aligned} \tag{4.29}$$

У четвртном случају, сви улазни битови мењају своје логичко стање. Није тешко закључити да ће у овом случају *carry* излаз свакако променити своје логичко

стање, без обзира на стање улаза. Према томе, транзициона активност c_{out} бита за овај случај је:

$$c_{out}^C = p \cdot q \cdot c_{in} \quad (4.30)$$

Коначни израз за транзициону активност c_{out} излаза добија се сабирањем транзиционих вероватноћа сва три случаја:

$$c_{out} = c_{out}^{A1} + c_{out}^{A2} + c_{out}^{B1} + c_{out}^{B2} + c_{out}^{B3} + c_{out}^C \quad (4.31)$$

Вероватноће да је c_{out} излаз ћелије потпуног сабирача у стањима логичке '0' или '1' рачунају се на следећи начин:

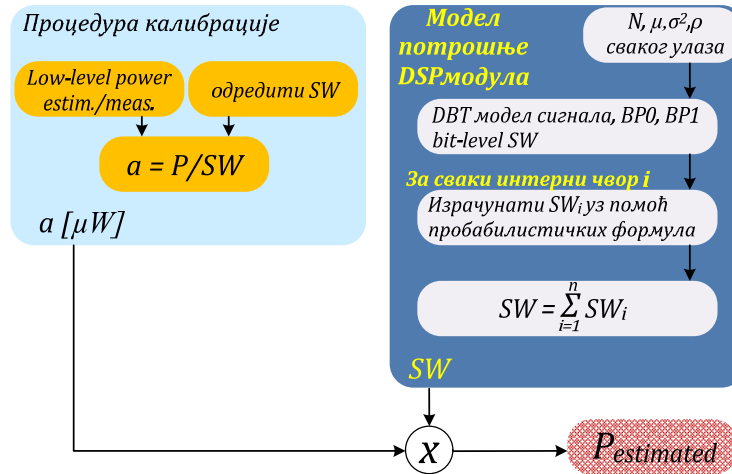
$$\begin{aligned} c_{out}^0 &= p^0 \cdot q^0 + c_{in}^0 \cdot (p^0 \cdot q^1 + p^1 \cdot q^0) \\ c_{out}^1 &= 1 - c_{out}^0 \\ s^0 &= (p^0 \cdot q^0 + p^1 \cdot q^1) \cdot c_{in}^0 + (p^0 \cdot q^1 + p^1 \cdot q^0) \cdot c_{in}^1 \\ s^1 &= 1 - s^0 \end{aligned} \quad (4.32)$$

Сада, једина преостала променљива потребна за прорачунавање транзиционе активности ћелије потпуног сабирача је транзициона вероватноћа излаза s . Ова променљива рачуна се на следећи начин:

$$s = (p \cdot q + (1 - p) \cdot (1 - q)) \cdot c_{in} + (p \cdot (1 - q) + q \cdot (1 - p)) \cdot (1 - c_{in}) \quad (4.33)$$

Ова методологија слична је методологији описаној у [CRP94]. Ипак, у [CRP94] метод је примењен на цео модул. Пошто рачунање вероватноће сигнала спада у групу NP-тешких проблема, [CRP94] за велика кола користе алгоритам поделе који ограничава број улаза у модул. Приступ коришћен у овој дисертацији примењује се на рачунање транзиционих активности неких основних ћелија DSP компоненти (бинарни сабирачи/одузимачи, множачи, делитељи, мултиплексери, регистри). До коначне транзиционе активности долази се пропагацијом транзиционе активности унутар модула. Последица тога су релативно прости изрази за укупну транзициону активност модула, без обзира на његову величину. Недостатак оваквог приступа је тај што не узима у обзир просторне зависности сигнала (нпр. када се исти улазни сигнал води у две или више различитих компоненти).

Графички приказ тока процене динамичке потрошње неког DSP модула приказан је на слици 4.6. За сваки модул потребно је најпре у кратком поступку емпиријске калибрације (детаљније описан касније) одредити константу a . Са друге стране, уз помоћ модела за процену потрошње DSP модула одређују се транзиционе активности у сваком чвору логичке структуре модула. Збир транзиционих активности свих интерних чворова представља укупну процењену транзициону активност модула која помножена са константом a даје његову процењену динамичку потрошњу. У поступку добијања модела за процену потрошње модула полази се од улазних сигнала. На основу њихових статистичких параметара одређују се транзиционе активности сваког бита (DBT модел сигнала). Након тога се за сваки интерни чвор, на основу табела истинитости модула, одређују пробабилистичке формуле за прорачун транзиционе активности чвора.



Слика 4.6: Графички приказ тока процене динамичке потрошње DSP модула

У [Jev09] читалац може наћи опис тока прорачуна и аналитичке изразе за транзиционе активности бинарних множача типа *RCA (Ripple Carry Array)*, *row adder tree multipliers* и *booth multipliers* (ова архитектура се користи за уграђене (*embedded*) множаче), као и *ripple-carry* бинарних сабирача. Поред тога, извршена је и анализа процене тачности модела, поређењем са измереним вредностима потрошње као и са процењеним вредностима добијеним од стране XPower алата. Евалуација је вршена за различите дужине (број битова) операнада као за различите статистике улазних Гаусових сигнала.

Каснијим анализама увидело се да се бинарни множачи понашају као нелинеарне компоненте тј. да на излазима множача део битова из LSB региона DBT модела сигнала (види слику 4.1) испољава транзициону активност мању од 0.5. Један од циљева ове дисертације био је да DBT модел сигнала замени новим моделом који ће, тиме што ће узети у обзир описану појаву на излазима множача, побољшати тачност процене транзиционе активности дизајна који садржи бинарне множаче, а самим тим и тачност процене динамичке потрошње. О овом моделу као и о току прорачуна транзиционе активности бинарних делитеља биће речи у наставку.

4.3.3. Гличеви

Као што је већ поменуто (види слику 2.19), гличеви су нежељене транзиционе активности у чворовима дизајна узроковане неизбалансираним путањама тј. различитим кашњењима сигнала који улазе у неку логичку компоненту. О техникама за минимизацију гличева као и о методологији за процену транзиционе активности која је последица гличева заинтересовани читалац може прочитати у [Jev09]. Транзициона активност која је последица гличева у [Jev09] моделована је следећом једначином:

$$G = k \cdot \sum_{i=1}^4 (1 - \rho_{1i}) \cdot (1 - \rho_{2i}) \cdot FA_i = k \cdot G' \quad (4.34)$$

где је G количина гличева, k емпиријски добијена константа која представља просечно гличовање на излазу LUT табеле, ρ_{1i} и ρ_{2i} су аутокорељациони

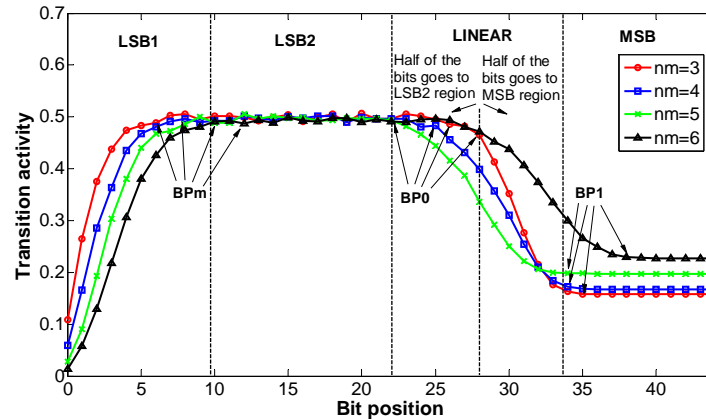
коефицијенти улаза на нивоу бита, а FA_i је број ћелија потпуног сабирача. Моделовање утицаја гличева на процену потрошње електричне енергије у овој дисертацији није разматрано.

4.4. ТВТ модел улазног сигнала

Као често присутан елемент у стазама података многих DSP дизајна бинарни множач конзумира значајан део укупне потрошње електричне енергије. Поред тога бинарни множач представља и главни извор нелинеарности у смислу да се својства (расподела) сигнала на излазу множача значајно мењају у односу на својства сигнала на његовим улазима.

Ако, на пример, посматрамо бинарни сабирач и за његове улазне сигнале (сабирке) претпоставимо да су стационарни и са Гаусовом расподелом (што је исправна претпоставка за сигнале у DSP апликацијама [LR95]) можемо бити сигурни да ће и излазни сигнал сабирача имати иста својства. Другим речима, сабирање два сигнала са Гаусовом расподелом за последицу ће имати трећи сигнал који ће такође имати Гаусову расподелу. Према томе, сигнале бинарног сабирача (и улазе и излаз) могуће је моделовати претходно описаним DBT моделом који, у зависности од статистике сигнала (средње вредности, варијансе и аутокорелације), разликује три региона транзиционе активности на нивоу бита (види слику 4.1).

Са бинарним множачима то није случај. У [BN02] примећено је да излаз множача има расподелу која је симетрична око средње вредности сигнала али ипак није Гаусова. Битови најмање тежине на излазу множача имају транзициону активност мању од 0.5 будући да једино два непарна броја за производ имају непаран број. Кад би се излаз множача моделовао DBT моделом сигнала процене потрошње дизајна који садржи множаче биле би мање тачне. Грешка процене потрошње била би још већа ако би дизајн садржао ланце множача што је случај са хардверским имплементацијама многих аритметичких функција (експоненцијалних, логаритамских, реципрочних [BMT+07]), FIR (Finite Impulse Response) и IIR (Infinite Impulse Response) филтара, FFT (Fast Fourier Transform) трансформација итд. Имајући то у виду, да би се побољшала ефикасност оптимизације потрошње оваквих имплементација, у дисертацији је представљен ТВТ (Triple-Bit Type) модел сигнала који уз помоћ статистике сигнала (средња вредност, варијанса, аутокорелација) и информације о броју претходно обављених множења у дизајну обезбеђује тачно рачунање транзиционе активности сваког бита на излазу множача. ТВТ модел допуњује већ постојећи DBT модел тиме што уводи нови регион транзиционе активности (LSB1 региона на слици 4.7). Овај нови регион обухвата LSB битове и расте са бројем претходно обављених множења, nm (број множења који се обавио у дизајну почев од улаза па до излаза множача чија се транзициона активност рачуна). Број битова који се налазе у LSB1 региону тј. нова граница региона BPm једнак је $BPm=2 \cdot nm$. Њихова транзициона активност у [BN02] рачуна се преко ентропије сигнала. Иако је апроксимација транзиционих активности битова у овом региону прилично тачна, транзициона активност осталих битова на излазу множача применом ентропије превазилази максималне дозвољене вредности. Стога је у овој дисертацији представљен нови начин за рачунање транзиционе активности у LSB1 региону.



Слика 4.7: Транзициона активност на нивоу бита у зависности од његове позиције за ланце различите дужине

Предложена је апроксимација транзиционе активности на нивоу бита која је представљена следећом формулом:

$$sw(i) = 0.5 - (0.5 - sw(0)) \cdot e^{-(0.25 + 2^{-(nm+1.25)}) \cdot i} \quad (4.35)$$

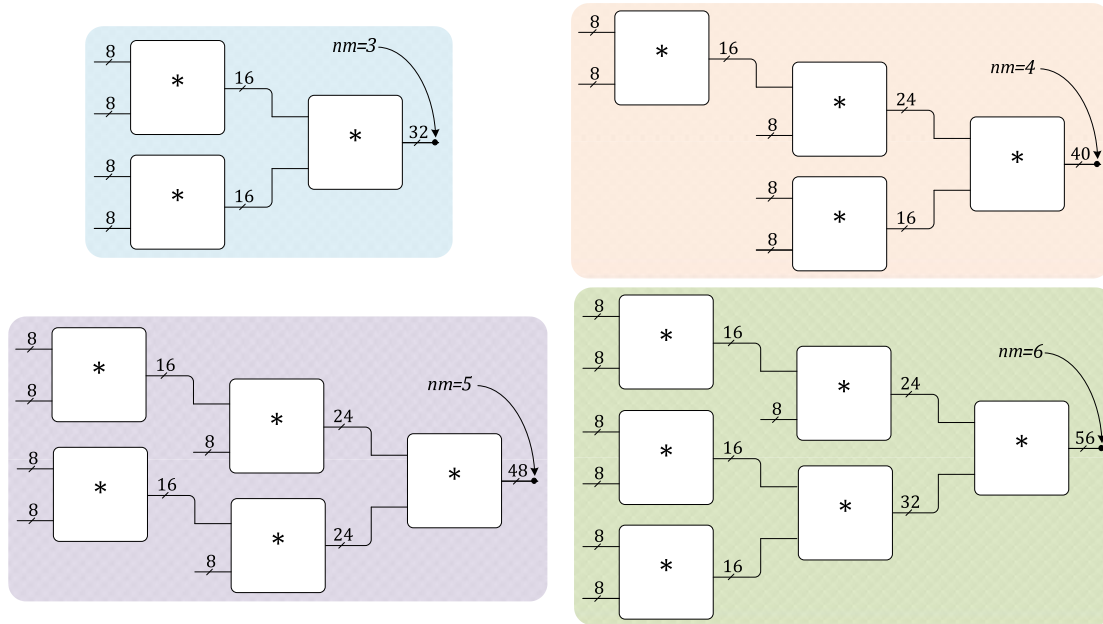
где је i позиција бита у оквиру дигиталне речи сигнала, nm број претходно обављених множења у дизајну а $sw(0)$ транзициона активност LSB бита која се рачуна према формули предложеној у [BN02]:

$$sw(0) = -(2^{-nm} \cdot \log_2(2^{-nm})) + (1 - 2^{-nm}) \cdot \log_2(1 - 2^{-nm}) \quad (4.36)$$

Транзиционе активности осталих бита као и границе региона $BP0$ и $BP1$ које деле LSB2, MSB и линеарни регион рачунају се исто као и у случају DBT модела сигнала.

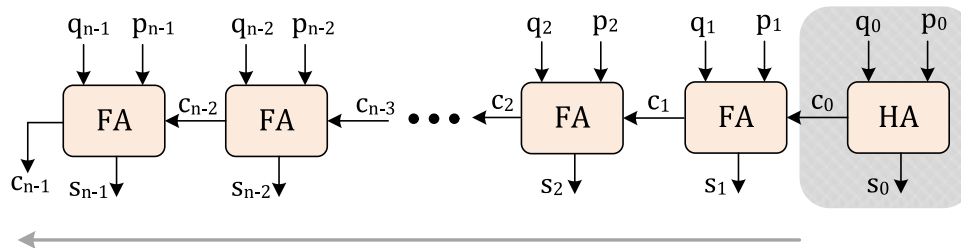
4.4.1. Евалуација ТВТ модела

Ради процене ефикасности ТВТ модела сигнала најпре смо, на примеру релативно простијих DSP дизајна приказаних на слици 4.8, поредили стварне вредности транзиционе активности унутар дизајна са процењеним вредностима добијеним применом DBT и ТВТ модела. У MATLAB програмском пакету извршено је генерисање сигнала са Гаусовом расподелом и 10000 вредности. Ови сигнали представљали су улазе у DSP дизајне чија се транзициона активност (стварна, DBT и ТВТ) такође рачунала у MATLAB-у. MATLAB функција за рачунање ТВТ транзиционе активности структуре множача поред статистике улазних сигнала (варијанса, аутокорелација, средња вредност) и броја њихових бита (N, M) у обзир узима и параметар nm – $sw_{TBT}(N, nm, var_x, rox, mix, M, nmy, vary, roy, miy)$. Бинарни множачи су RCA типа. Примена модела за рачунање транзиционе активности састоји се у томе да се крене од улаза компоненте и да се врши пропација рачунања транзиционих активности и вероватноћа транзиција кроз компоненту, све док се не обраде све логичке структуре у оквиру компоненте [JC10]. Овај поступак илустрован је на примеру сабирача *ripple-carry* структуре приказане на слици 4.9.



Слика 4.8: DSP дизајни са различитим бројем множача за евалуацију ТВТ модела сигнала

Приликом рачунања транзиционе активности *ripple-carry* сабирача креће се од крајње леве (осенчане) ћелије полусабирача и рачунају се транзиционе активности и вероватноће транзиција излазног (s_0) и *carry* (c_0) бита, и то помоћу израза који се добијају на основу табеле истинитости полусабирача, поступком објешњеним у поглављу 4.2.2. Логички изрази за ћелију полусабирача могу се наћи у додатку А. Након тога приступа се следећој логичкој компоненти у ланцу (потпуном сабирачу), све док се не дође до крајње десне ћелије у ланцу.



Слика 4.9: Демонстрација методе рачунања транзиционе активности на *ripple-carry* сабирач

Пошто су нам сада познате транзиционе активности свих логичких компоненти у структури *ripple-carry* сабирача њиховим сабирањем добија се укупна процењена транзициона активност структуре:

$$SWest = \sum_{i=0}^{n-1} (\alpha_i^c + \alpha_i^s) \quad (4.37)$$

где је α_i транзициона активност i -те ћелије у низу.

Процена укупне транзиционе активности DSP структура са слике 4.8 одвија се на следећи начин:

1. Најпре се, на основу статистике улазних сигнала (средња вредност, аутокорељација, варијанса) рачунају границе региона $BP0$, $BP1$ и $BP2$ (према

једначинама 4.21, 4.19 и 4.24, респективно). За ТВТ модел сигнала се, на основу параметра nm рачуна и вредност границе BPm .

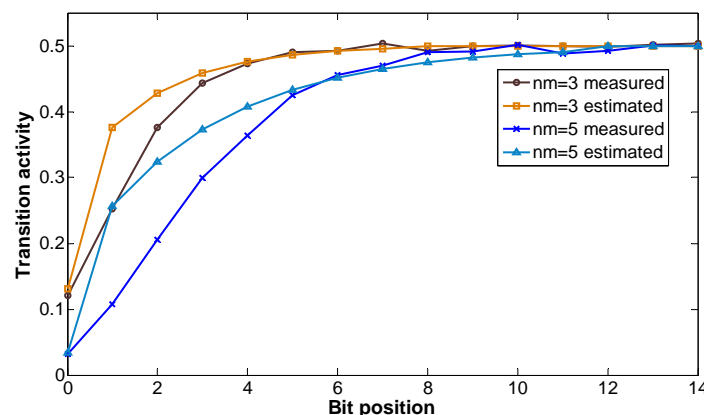
2. Транзиционе активност улазних битова као и њихове вероватноће транзиције рачунају се према једначинама 4.17 и 4.18, респективно. За ТВТ модел, транзиционе активности битова најмање тежине (од LSB бита до границе BPm) рачунају се према формули 4.35.
3. Након тога се рачунају транзициона активности и вероватноће транзиције основних ћелија у структури RCA множача, по формулама датим у [Jev09], на начин илустрован на слици 4.9.
4. Сумирањем парцијалних транзиционих активности (са свих чворова RCA структуре) добија се процењена вредност транзиционе активности множача.
5. Кораци 1 – 4 понављају се за сваки множач из DSP дизајна. Сабирањем транзиционих активности свих множача добија се укупна процењена транзициона активност DSP структуре.

Релативне грешке оба модела (ДБТ и ТВТ) у односу на стварне вредности транзиционих активности DSP дизајна приказане су у табели 4.1. Очигледно је да моделовање улазних сигнала бинарних множача ТВТ моделом врло повољно утиче на тачност процене транзиционе активности.

Табела 4.1: Релативне грешке између суме стварних и процењених транзиционих активности

nm	3	4	5	6
ErrDBT[%]	4.20	6.29	6.78	10.48
ErrTBT[%]	0.87	1.07	-0.17	2.08

На слици 4.10 приказане су стварне и ТВТ моделом процењене (апроксимирани једначином 4.35) транзиционе активности на нивоу бита у BPm региону за DSP структуре са слике 4.8 код којих је $nm=3$ и $nm=5$.



Слика 4.10: Стварна (BLT) у односу на процењену (ТБТ) транзициону активност

Након овога, ТВТ модел примењен је на мало компликованије дизајне. С једне стране, сваки од три тестна DSP дизајна имплементиран је на Virtex II Pro XC2VP30 FPGA чип (XUP развојне плоче), а потрошња дизајна мерена је на начин описан у глави 3 и додатку А. С друге стране, на исте DSP дизајне примењени су методи за процену транзиционе активности логичких компоненти тј. методи за процену

динамичке потрошње логике. Транзициона активност процењивана је применом TBT и DBT модела. Такође је, као референца узимана и стварна вредност транзиционе активности (BLT).

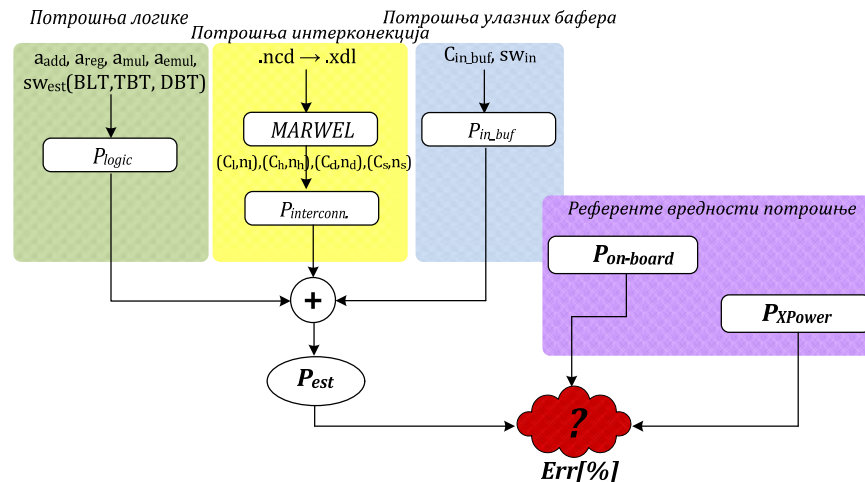
Све вредности транзиционих активности добијене су у MATLAB-у. Динамичка потрошња сваке логичке компоненте добија се множењем њене транзиционе активности емпиријски добијеном константом a (види једначину 4.1). Вредности ове константе за сабираче множаче (LUT и embedded) и регистре (који улазе у састав DSP дизајна) дате су у табели 4.2. Емпиријски поступак добијања константе a биће објашњен касније, на примеру бинарног делитеља.

Табела 4.2: Вредности емпиријске константе a за имплементацију на XC2VP30 FPGA чип

	Reg	Add	Mul	Emb mul
$a[\mu\text{W}]$	258.3	79.3	132.5	293.8

Измерене и моделима процењене вредности динамичке потрошње логике се након тога пореде да би се добила релативна грешка модела за процену. Поред измерене вредности за референцу је узета и вредност потрошње добијена помоћу XPower алата из ISE 10.1 верзије софтвера.

Ток евалуације модела илустрован је на слици 4.11. Кад се након примене MARWEL алата добије број линија сваке врсте које се у дизајну користе да повежу улазе и излазе са логиком (глобалне интерконеције), као и логичке блокове између себе (локалне интерконеције) знајући ефективну капацитивност сваке линије (види табелу 3.1) као и транзициону активност на њој (добија се уз помоћ MATLAB-а) прорачунава се динамичка потрошња интерконеција (види једначину 3.2). Динамичка потрошња улазних бафера рачуна се на основу ефективне капацитивности улазних бафера (3.52pF) и транзиционе активности улазних сигнала. Сабирањем динамичке потрошње интерконеција и улазних бафера са процењеном динамичком потрошњом логике добија се вредност динамичке потрошње која се пореди са измереном вредношћу и XPower референцом.



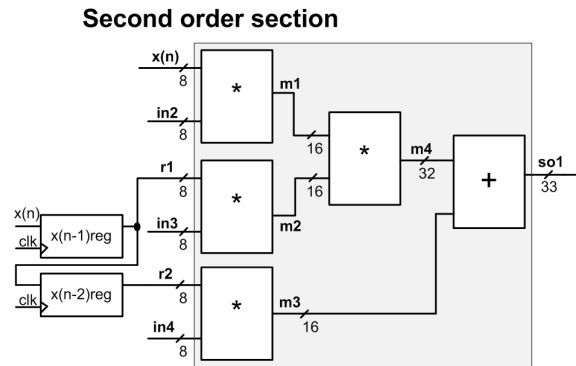
Слика 4.11: Ток евалуације TBT модела

Што се тиче DSP дизајна који су коришћени за евалуацију, прва два су релативно мала и одговарају имплементацији следећих логичких функција:

$$\begin{aligned}
 DSP_1 &= (x_2 \cdot x_3) \cdot x_2 + (x_1 + x_3) \cdot x_2 \\
 DSP_2 &= ((x_1 + x_2) \cdot (x_3 + x_4) + x_1 \cdot x_2) \cdot x_2 \cdot (x_3 + x_4)
 \end{aligned}
 \tag{4.38}$$

DSP₃ дизајн је прилично већи и има структуру која подсећа на структуру FIR филтра 16. реда имплементирану као каскадна реализација осам секција другог реда. Једна секција DSP₃ дизајна приказана је на слици 4.12.

Три множача из секције имплементирани су у LUT табеле док је један множач *embedded* типа.



Слика 4.12: Једна од осам каскасно везаних секција DSP₃ дизајна

Сва три DSP дизајна су синхронa. Тактна фреквенција прва два дизајна је 50MHz. Имајући у виду комплексност DSP₃ дизајна тактна фреквенција му је смањена на 16MHz, да би се избегло евентуално повећање статичке потрошње (услед рада на већој фреквенцији тј. услед веће динамичке потрошње могућа су загревања чипа а самим тим и повећања статичке потрошње која је зависна од температуре). Табела 4.3 приказује резултате поређења за сваки дизајн, за различите вредности аутокорељације ρ улазних сигнала.

Табела 4.3: Релативне грешке процене потрошње за три модела сигнала (BLT, TBT, DBT) и XPower (XPw) алат, за различите вредности аутокорељације улазних сигнала

Benchmark	Slices	Emb. mult.	Comp. time[s]	ρ	Err(BLT) [%]	Err(TBT) [%]	Err(DBT) [%]	Err(XPw) [%]
DSP1	212	2	0.92	0	10.03	7.6	17.48	328.79
				0.9	6.94	1.49	6.62	316.48
				0.99	9.33	7.6	9.59	281.70
				0.9995	9.49	8.71	13.68	246.91
DSP2	192	2	1.1	0	7.92	4.5	11.58	258.45
				0.9	7.51	1.03	7.6	233.50
				0.99	12.18	7.81	11.54	216.23
				0.9995	22.24	21.24	30.99	245.27
DSP3	2977	8	91.95	0	-0.38	9.91	38.3	455.08
				0.9	-1.55	8.99	37.38	455.06
				0.99	-1.56	10.05	38.24	437.46
				0.9995	-0.78	15.25	41.27	442.62

Прве две колоне показују број заузетих слајсева сваког DSP дизајна као и број *embedded* множача коришћених у дизајну. Времена прорачуна процене потрошње која су у следећој колони праћена су коефицијентима аутокорељације и релативним грешкама сваког модела у односу на измерену вредност динамичке потрошње. Комплексност DSP₃ дизајна потврђена је бројем заузетих слајсева као и временом потребним за процену динамичке потрошње електричне енергије. Што се тиче релативних грешака, може се закључити да TBT модел даје боље

процене потрошње (четири до пет пута) у односу на DBT модел, за све DSP дизајне и за све коефицијенте аутокорелације. Даље, TBT модел постиже највеће побољшање у односу на DBT модел у случају DSP₃ дизајна. Ово се може објаснити чињеницом да се DSP₁ и DSP₂ дизајни састоје од по неколико сабирача и множача и да имају свега неколико битова у LSB1 региону тако да је ефекат замене DBT модела TBT моделом једва приметан. Број сабирача и множача у DSP₃ дизајну је много већи што за последицу има значајно повећање броја LSB1 битова тако да је утицај тачнијег модела сигнала (TBT) доста очигледнији. Из табеле се такође може видети да XPower алат није у стању да обезбеди тачне процене динамичке потрошње дизајна.

4.5. Модел за процену потрошње бинарног делитеља

У овом поглављу биће представљена методологија процене динамичке потрошње електричне енергије бинарних делитеља имплементираних на FPGA у облику IP језгра. Методологија узима у обзир структуру делитеља и статистике улазних сигнала делитеља и дељеника (средњу вредност, варијансу, аутокорелацију). За процену укупне транзиционе активности у структури делитеља коришћен је аналитички модел, а велика међусобна зависност података која је примећена на улазима градивних елемената делитеља моделована је на одговарајући начин како би се побољшала тачност модела. Методологија је у стању да обезбеди брзе и тачне процене потрошње. За евалуацију модела коришћена су мерења потрошње IP соге делитеља имплементираних на XUP развојној плочи као и процене потрошње добијене XPower алатом.

4.5.1. Бинарни делитељи

За многе DSP апликације дељење представља есенцијалну операцију (дигитално процесирање сигнала, дигитална обрада слике [GSS08], Levinson-Durbinov алгоритам у LPC кодеру говора [GG04], контрола робота итд.). У исто време, ако са становишта хардверске имплементације поредимо четири основне аритметичке операције имплементација бинарног делитеља је без сумње најкомплекснија (у погледу потрошње електричне енергије, брзине рада и заузећа површине чипа). Стога је потреба за моделом за брзу и тачну процену потрошње бинарног делитеља очигледна. Пре него што се фокусирамо на модел за процену транзиционе активности структуре бинарног делитеља треба рећи неколико речи о самим бинарним делитељима.

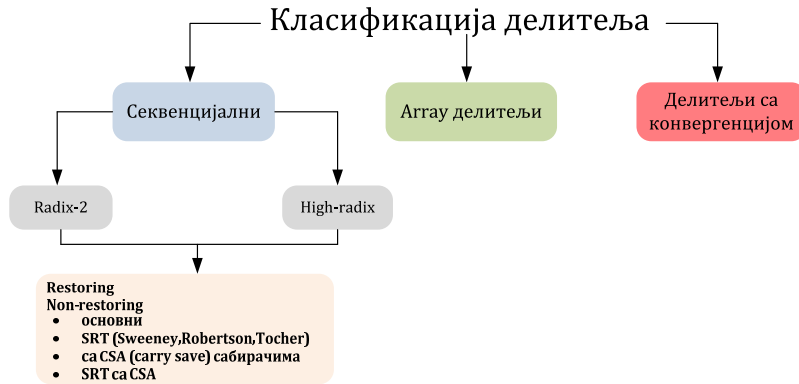
Операнди дељења могу бити означени или неозначени бројеви, целобројни или разломљени, са покретним или непокретним зарезом. Једначине 4.39 и 4.40 представљају основне једначине дељења:

$$X = Y \cdot Qt + Rem \quad (4.39)$$

$$X = Y \cdot Qt.Re m \quad (4.40)$$

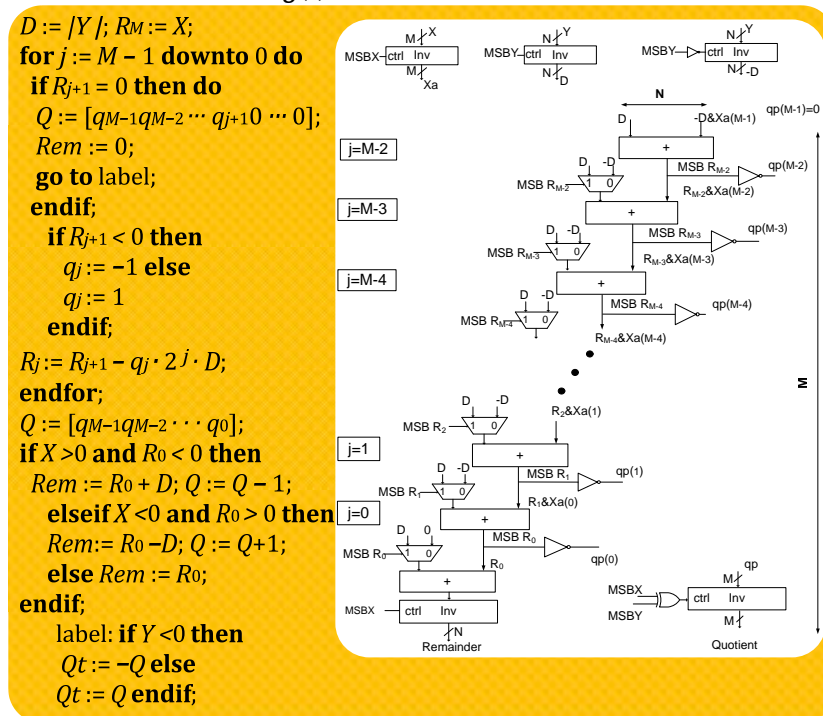
Дељењем броја X бројем Y добијамо количник Qt и остатак дељења Rem . У једначини 4.39 остатак дељења је цео број мањи од Y док је у једначини 4.40

остатак број мањи од 1. Ако је број X представљен са M битова а број Y са N битова, за представљање количника Qt у најгорем случају биће потребно M битова а за остатак Rem N битова. На слици 4.13 приказана је класификација основних алгоритама бинарног дељења [ТКТ05].



Слика 4.13: Класификација алгоритама бинарног дељења

Radix-2 алгоритми генеришу по један бит количника и остатка у секвенци (тактном циклусу) док *High-radix* алгоритми генеришу по више битова. Алгоритам бинарног дељења који се користи у Xilinx-овим IP језгрима је *Radix-2 non-restoring* алгоритам те му је стога потребно посветити више пажње. На слици 4.14 приказана је програмска секвенца дељења заједно са хардверском структуром *Radix-2 non-restoring* делитеља.



Слика 4.14: *Radix-2 non-restoring* алгоритам бинарног дељења

Имајући, у општем случају, на улазу дељеник X и делилац Y као M - и N -то битне целе означене бројеве представљене у другом комплементу, респективно, алгоритам дељења на излазу делитеља даје количник Qt и остатак Rem . Количник

и остатак су такође M - и N -то битни цели означени бројеви представљени у другом комплементу, тако да је испуњена једна од једначина 4.39 или 4.40. Остатак Rem је истог знака као и дељеник X и може бити *non-fractional* (тј. цео број, $IntRmd$, према једначини 4.39) и *fractional* типа (број мањи од 1, $FracRmd$, према једначини 4.40). Остатак типа *fractional* добија се још једним дељењем целобројног остатка делиоцем Y . Једначина 4.41 представља везу између N -то битног, означеног целобројног и *fractional* остатка.

$$FracRmd = \frac{IntRmd \cdot 2^{(N-1)}}{Y} \quad (4.41)$$

Са сваким проласком кроз *for* петљу генерише се по један бит количника q_j (и то почев од бита највеће тежине). У случају да је остатак R_j једнак нули кроз *for* петљу пролази се мање од M пута. j -ти пролазак кроз *for* петљу даје вредност привременог остатка R_j као и бит количника q_j који зависи од привременог остатка R_{j+1} добијеног у претходној итерацији.

Што се тиче хардверске структуре, на почетку се, уз помоћ инвертора са контролним улазом *ctrl*, одређују апсолутне вредности дељеника и делиоца, X и D , као и негативне вредности делиоца, $-D$. Ове вредности затим улазе у ланац сабирача и мултиплексера. Број сабирача и мултиплексера у ланцу одређен је бројем битова дељеника (M), док њихова ширина (број битова на улазу) зависи од броја битова делиоца (N). Бит највеће тежине (MSB) на излазу сваког сабирача одређује по један бит апсолутне вредности количника као и селекторски улаз мултиплексера. Излаз из последњег сабирача представља целобројни остатак дељења. На крају се врше корекције знакова количника и остатка у зависности од комбинације знакове дељеника и делиоца.

4.5.2. Преглед досадашњих активности

У литератури је могуће наћи неколико радова орјентисаних ка хардверској имплементацији алгоритама бинарног дељења [Sor06, EM09, SD09]. Међутим, главни циљеви ових радова су углавном минимизација потрошње [EM09] и/или кашњења [Sor06, SD09].

У [EM09] аутори имплементирају *radix-10* алгоритам дељења на Virtex 5 FPGA чип уз помоћ множача ограничене прецизности, сабирача и табела са меморисаним стањима. Иако је смањен проценат заузећа FPGA ресурса забележено је повећање кашњење на везама што се негативно одражава на брзину рада.

У [SD09] представљен је алгоритам који смањује време за које се обавља операција бинарног дељења тиме што се уместо *radix-2* алгоритма користи *radix-2ⁱ* (*high-radix*) алгоритам. Овим се омогућава да се у једном тактном интервалу добије више битова количника, а самим тим и мања латенција. Ипак цена мање латенција плаћена је смањеном максималном радном фреквенцијом у односу на Xilinx IP core делитеље.

Врло мали број радова бави се потрошњом електричне енергије бинарних делитеља. У [SDB+04] вршено је, заједно са кашњењем и заузећем површине, поређење потрошње бинарних делитеља различитих архитектура. Најбоља архитектура са становишта површине-потрошње-кашњења била је архитектура

базирана на *SRT radix-2* алгоритму. Овај алгоритам сличан је *non-restoring* методу. Једина разлика је у употреби табеле са меморисаним стањима за добијање битова коефицијента.

Ипак, ниједан од поменутих радова не бави се проценом потрошње бинарних делитеља имплементираних на FPGA компоненте. Неколико модела за процену потрошње на нивоу гејтова (низак ниво апстракције система) FPGA чипа представљено је у [PYW02, AN04]. Ове моделе могуће је користити за процену потрошње било које логичке компоненте на врло ниском нивоу апстракције. Међутим, велики недостатак ових модела је изузетно велико време потребно за процену потрошње. Неколико бржих метода на високом нивоу апстракције система за процену потрошње сабирача и множача предложено је у [ЛТВО4, CGC05, JCC08, Јев09, ЈС10]. Међутим, по информацијама које су нама на располагању, не постоји рад који се фокусира на процену потрошње бинарних делитеља на високом нивоу апстракције система.

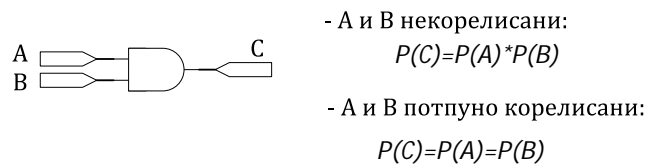
4.5.3. Рачунање транзиционе активности унутар делитеља

Модел за процену транзиционе активности (динамичке потрошње) структуре *Radix-2 non-restoring* бинарног делитеља (користи се за имплементацију Xilinx IP језгра делитеља) представљен је у [ЈС11]. Базира се на специфичној хардверској структури делитеља (приказаној на слици 4.14) која се састоји од ланца сабирача и мултиплексера. На основу броја битова и статистике улазних сигнала дељеника и делиоца, поступком објашњеним у поглављу 4.3.1 (види слику 4.9), врши се пропација транзиционе активности и вероватноће транзиције кроз структуру бинарног делитеља. Пошто су аналитички изрази за транзициону активности и вероватноћу транзиције бинарних сабирача добијени у [Јев09] остаје нам само да, применом пробабилистичке методологије добијемо аналитичке изразе за инверторе и мултиплексере који, поред сабирача, улазе у састав хардверске структуре *Radix-2 non-restoring* бинарног делитеља. Применом пробабилистичке методологије описане у поглављу 4.2.2, уз помоћ табела истинитости и логичких структура инвертора и мултиплексера добијени су аналитички изрази за транзиционе активности и вероватноће транзиција излазних и *carry* битова у структурама инвертора и мултиплексера. Ови изрази могу се пронаћи у додатку А. Имајући на располагању аналитичке изразе за рачунање транзиционих активности свих логичких компоненти које улазе у састав бинарног делитеља проблем рачунања укупне транзиционе активности чини се једноставним.

Међутим, за разлику од бинарних сабирача и множача код којих се није јављао ефекат међусобне корелације, због специфичне логичке структуре бинарног делитеља улази у сабираче у ланцу су међусобно корелисани, чак иако глобални улази, дељеник и делитељ, нису уопште међусобно корелисани. Ако детаљније анализирамо слику 4.14 можемо, на пример, видети да су два улаза инвертора међусобно корелисана и то са максималним коефицијентом међусобне корелације (1 или -1), будући да је контролни бит инвертора (*ctrl*) једнак (или инвертован) биту највеће тежине дељеника или делиоца. Улази мултиплексера такође су јако зависни (*D* и *-D*, коефицијент међусобне корелације једнак -1) а постоји и јака корелација (од приближно -1) између улаза првог сабирача, на врху

ланца, будући да су ови једнаки апсолутној вредности делиоца (D) и његовој негативној вредности ($-D$) помереној за једну позицију у лево. Међусобна корелација са опадајућим трендом присутна је и на улазима осталих сабирача у ланцу.

У сваком случају, претходно описани пробабилистички метод није у потпуности применљив у присуству међусобне корелације на улазима логичких компоненти. Емпиријски је утврђено да у овом случају постоје прецењивања транзиционе активности (процењене вредности транзиционе активности су веће од правих), нарочито у MSB региону на излазима бинарних сабирача. Утицај међусобне корелације улаза може се објаснити на следећем примеру: посматрајмо једно двоулазно логичко I коло приказано на слици 4.15. Претпоставимо, најпре, да су улази A и B тотално независни један од другог (коефицијент међусобне корелације једнак 0). У том случају, вероватноћа да је излаз C у стању логичке '1' једнака је производу вероватноћа да су оба улаза у стању '1'. Ако сада претпоставимо други, екстремни, случај да су улази A и B са максималним коефицијентом међусобне корелације тј. да су једнаки, вероватноћа да је излаз C у стању '1' није више једнака производу вероватноћа улаза (види слику 4.15). Исто се дешава и са транзиционом активношћу па је стога потребна мала модификација у приступу рачунања транзиционе активности тј. узимање у обзир међусобне зависности улаза. Поступак рачунања транзиционе активности структуре бинарног делитеља, уз претпоставку да је средња вредност улазних сигнала једнака нули, описан је у наставку. Улазни сигнали моделовани су DBT моделом сигнала.



Слика 4.15: Утицај међусобне корелације улаза на вероватноћу/транзициону активност излаза

Инвертори се користе за рачунање апсолутних и негативних вредности дељеника или делиоца и контролисани су њиховим битовима знака (види слику 4.14). Услед присуства међусобне корелације на улазима инвертора најпре се према једначинама 4.21 и 4.19 рачунају границе $BP0$ и $BP1$ излаза инвертора. У случају одређивања апсолутне вредности дељеника X_a , у изразима за одређивање граница, варијанса σ и аутокорелација ρ једнаке су варијанси и аутокорелацији дељеника X , док се приликом одређивања апсолутне и негативне вредности делиоца узимају варијанса и аутокорелација делиоца Y . Након тога пробабилистички метод примењује се за рачунање транзиционих активности и вероватноћа транзиција свих битова почев од LSB па до границе $BP1$. Транзициона активност MSB битова свих инвертора једнака је нули. Ово је очигледно имајући у виду да су битови у овом региону константни и једнаки '0' у случају одређивања апсолутне вредности, или константни и једнаки '1' у случају рачунања негативне вредности. Будући да су константни ови битови не мењају своје логичко стање па самим тим нема ни транзиционе активности.

У случају да инвертор одређује апсолутне вредности (X_a или D) вероватноћа да су битови у MSB региону у стању '1' једнака је 0, док је у случају одређивања негативне вредности делиоца ($-D$) једнака 1. Вероватноћа да су битови у LSB

региону у стању '1', према DBT моделу сигнала једнака је 0.5. Вероватноћа у линеарном региону представља линеарну апроксимацију између вероватноћа у LSB и MSB области. Аналитички изрази за транзиционе активности и вероватноће транзиција ћелија инвертора могу се наћи у додатку А.

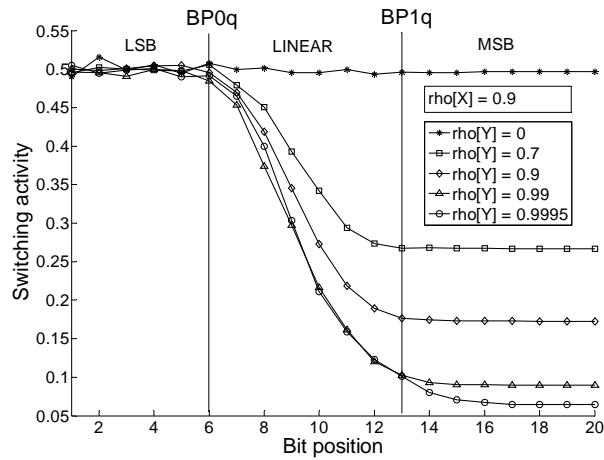
Упркос присуства међусобне корелације на улазима мултиплексера коришћен је пробабилитички метод за рачунања транзиционе активности и вероватноће транзиције излаза мултиплексера. Емпиријски је утврђено да овакав приступ уноси веома малу грешку у процени транзиционе активности. Аналитички изрази за мултиплексер такође се налазе у додатку А.

Што се тиче ланца сабирача, статистике (средња вредност, варијанса, аутокорелација) улазних сигнала првог сабирача у ланцу (D и $-DX(M-1)$) одговарају статистици делиоца Y будући да су улази у сабирач једнаки апсолутној вредности делиоца и његовој негативној вредности помереној за једно место улево. Према томе, границе $BP0$ и $BP1$ излаза првог сабирача једнаке су границама делиоца. Даље, исте границе користе се и за излазе осталих сабирача у ланцу. Ова апроксимација доста упрошћава модел а такође је емпиријски утврђено да не уноси велику грешку у процену транзиционе активности. Према DBT моделу сигнала (види слику 4.1), једини преостали параметри који су нам потребни да одредимо транзиционе активности свих битова на излазу сабирача су транзиционе активности MSB битова као и вероватноће да су у стању '1'. Кад бисмо за рачунање транзиционе активности MSB битова излаза првог сабирача применили једначину 4.17 добили бисмо прецењене вредности за транзициону активност. Ова велика грешка у процени транзиционе активности последица је велике међусобне зависности улаза у први сабирач. Што је већа корелација између улаза испољава се мање транзиционе активности унутар логичке структуре. Међусобна зависност улаза у следеће сабираче у ланцу све је слабија и слабија али ипак није занемарива. Због тога је у наставку предложен нови модел за рачунање транзиционих активности MSB битова свих сабирача у ланцу.

Детаљнијом анализом структуре делитеља може се уочити да бит највеће тежине излаза сваког сабирача представља по један бит апсолутне вредности количника. Према томе, све што треба да урадимо је да нађемо транзициону активност на нивоу бита од количника и да заменимо транзициону активност MSB битова количника са нулом будући да нам треба транзициона активност апсолутне вредности количника. Транзициона активност сваког бита апсолутне вредности количника једнака је транзиционој активности MSB бита на излазу једног сабирача у ланцу.

На слици 4.16 приказана је транзициона активност на нивоу бита од количника дељења два сигнала са Гаусовом расподелом. Аутокорелација дељеника X једнака је 0.9 док аутокорелација делиоца варира од 0 до 0.9995. Са слике се лепо види да транзициона активност количника има облик транзиционе активности сигнала са Гаусовом расподелом тј. да јасно можемо разликовати три различита региона транзиционе активности: LSB регион, тзв. линеарни и MSB регион. Имајући то у виду, сигнал количника апроксимирали смо сигналом са Гаусовом расподелом и еквивалентном варијансом и аутокорелационим коефицијентом чији су изрази дати у наставку. Иако ова два еквивалентна параметра (варијанса и аутокорелација) не представљају праве параметре

сигнала делитеља примењени у једначинама 4.21 и 4.19 ипак дају задовољавајуће резултате за граничне вредности $BP0$ и $BP1$ битова количника.



Слика 4.16: Транзициона активност (на нивоу бита) количника дељења

Еквивалентна варијанса количника апроксимирана је са:

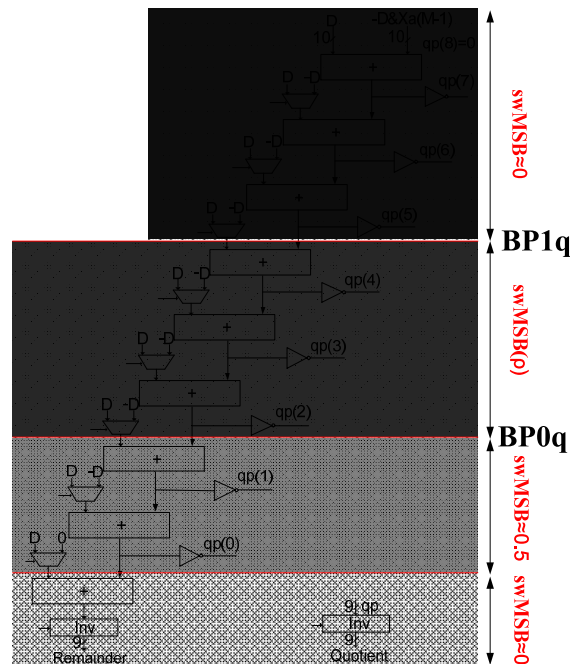
$$\sigma_Q = \frac{\sigma_X}{\sigma_Y} \quad (4.42)$$

За еквивалентан коефицијент аутокорелације сигнала количника емпиријским путем је утврђено да је једнак производу аутокорелационих коефицијената дељеника и делиоца:

$$\rho_Q = \rho_X \cdot \rho_Y \quad (4.43)$$

Са ове две вредности можемо израчунати граничне вредности $BP0$ и $BP1$ апсолутне вредности количника. Ове граничне вредности одређују број сабирача чија је транзициона активност MSB битова једнака 0.5 (LSB регион сигнала количника), број сабирача са транзиционом активношћу MSB битова једнаком 0 (MSB регион количника) и број сабирача чија је транзициона активност MSB битова линеарно апроксимирана између 0 и 0.5. Због прегледности, на слици 4.17 приказана је структура 9×9 делитеља на којој су сабирачи јасно подељени према транзиционим активностима њихових MSB битова. Излаз последњег сабирача представља остатак дељења који је, према алгоритму, увек позитиван. Самим тим, његови MSB битови имају транзициону активности једнаку нули.

За MSB битове на излазима сабирача вероватноће (p_i) да су у стању логичке '1' рачунају се према једначини 4.17, са транзиционим активностима MSB битова (sw_i) добијеним на управо описани начин. Ове вероватноће потребне су нам приликом рачунања транзиционе активности carry битова у структурама бинарних сабирача.



Слика 4.17: Подела на зоне структуре бинарног делитеља према транзиционим активностима MSB битова на излазима сабирача

Сумирано, процедура рачунања транзиционе активности структуре бинарног делитеља састоји се у следећем:

- 1) Прорачунати граничне вредности $BP0$ и $BP1$ инвертора. За добијање транзиционе активности излазних битова инвертора, почев од бита најмање тежине па до бита одређеног границом $BP1$ примењује се пробабилистички метод. Транзиционе активности MSB битова на излазу инвертора једнаке су нули. Сумирањем транзиционих активности свих битова на излазу инвертора добија се њихова укупна транзициона активност SW_{inv} . Вероватноће да су излази инвертора у стању логичке '1' рачунају се помоћу једначине 4.17.
- 2) Израчунати граничне вредности $BP0q$ и $BP1q$ апсолутне вредности количника. На основу њих одредити транзиционе активности битова у MSB зони на излазима бинарних сабирача (број сабирача чији MSB битови имају транзициону активност 0.5, број сабирача са нултом транзиционом активношћу у MSB зони и број сабирача са MSB транзиционом активношћу линеарно апроксимираном између 0 и 0.5). Вероватноће да су MSB битови сабирача у стању логичке '1' рачунају се према једначини 4.17.
- 3) Израчунати граничне вредности $BP0$ и $BP1$ ланца бинарних сабирача. Транзициона активност битова у LSB региону на излазима сабирача једнака је 0.5. Сваком сабирачу придружити одговарајући бит апсолутне вредности количника. Транзициона активност у MSB региону сабирача једнака је транзиционој активности одговарајућег бита количника. Извршти екстраполацију транзиционе активности битова у линеарном региону (линеарну апроксимацију између транзиционе активности у MSB и LSB области). Применити пробабилистички метод за рачунање транзиционе активности carry битова. Сабирањем транзициони активности у чворовима

свих сабирача добија се укупна транзициона активност ланца сабирача $SW_{addchain}$.

- 4) Применити пробабилистички метод за рачунање транзиционе активности ланца мултиплексера. Сума транзиционих активности свих битова на излазима мултиплесера једнака је SW_{mux} .
- 5) Укупна транзициона активност структуре бинарног делитеља, SW , једнака је суми $SW_{inv}+SW_{addchain}+SW_{mux}$.

Из претходно изложеног може се закључити да је модел за процену транзиционе активности параметризован и да су му за процену потребни следећи параметри: димензије (број битова) и статистика (аутокорелација и варијанса) улазних сигнала дељеника и делиоца.

Поред транзиционе активности, за процењивање потрошње електричне енергије, према изразу 4.1, неопходно нам је и познавање емпиријске константе a . Ова константа, будући непозната јер у себи садржи излазну капацитивност C_i , одређује се на следећи начин: на FPGA чип имплементира се IP core бинарног делитеља и мери се потрошња таквог дизајна и то на начин описан у глави 3. Након тога се рачуна транзициона активност такве структуре. Уз измерену потрошњу и познату транзициону активност лако је одредити константу a (види једначину 4.1). Са циљем да ова константа буде што тачније одређена врши се неколико мерења потрошње (за имплементирани делитеље различитих димензија и статистика сигнала) и неколико рачунања транзиционе активности. Након тога се, на основу више парова тј. тачака (P_i, sw_i) , у MATLAB програму одређује константа a која најбоље испуњава услов $P_i = a \cdot sw_i$, за $i=1$ до N , где је N број тачака (P_i, sw_i) . Када једном емпиријски одредимо константу a можемо је користити за било који други бинарни делитељ (са различитим димензијама и статистикама сигнала) имплементиран на исти FPGA чип. Све што треба да урадимо за процену потрошње је да одредимо укупну транзициону активност структуре делитеља, помножимо је са познатом константом a и добијемо процењену вредност динамичке потрошње енергије.

4.5.4. Евалуација модела за процену потрошње бинарног делитеља

Евалуацију модела за процену транзиционе активности (динамичке потрошње) бинарног делитеља поделили смо у два дела. У једном делу смо процењивали тачност рачунања транзиционе активности у структури делитеља (поређењем процењене вредности са стварним вредностима транзиционе активности) док је у другом делу евалуирана способност модела да процењује динамичку потрошњу електричне енергије (кроз поређење са измереним вредностима потрошње или вредностима добијеним уз помоћ XPower Analyzer алата за процену потрошње).

Евалуација модела за рачунање транзиционе активности у структури бинарног делитеља

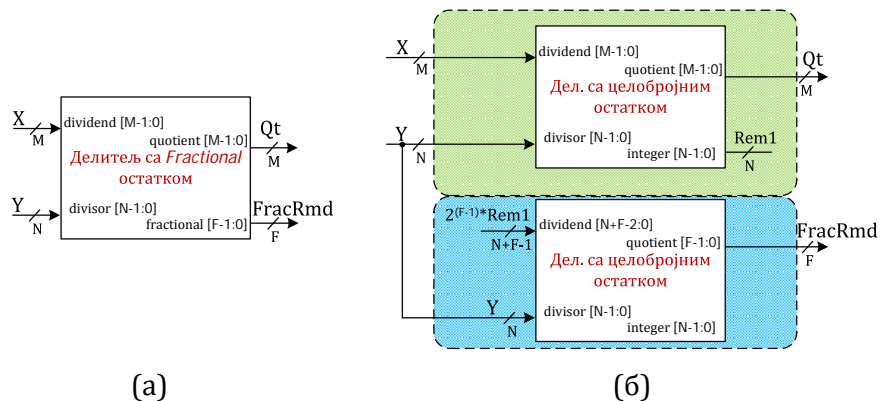
Приликом овог евалуационог сета коло бинарног делитеља описано је у MATLAB програмском пакету. За улазне сигнале дељеника и делиоца са Гаусовом

расподелом и различитим димензијама (бројем битова) рачунана је стварна вредност укупне транзиционе активности у свим чворовима логичке структуре делитеља (sw_{act}). Са друге стране, на основу броја битова улазних сигнала и њихове статистике (средња вредност, варијанса, аутокорелација), на начин описан у претходном поглављу, добијена је процењена вредност транзиционе активности структуре бинарног делитеља (sw_{est}). Након тога је рачуната релативна грешка процењене у односу на стварну вредност транзиционе активности:

$$Err[\%] = \frac{(sw_{est} - sw_{act})}{sw_{act}} \cdot 100 \quad (4.44)$$

У табелама 4.4 и 4.5, за делитеље са целобројним и *fractional* остатком, респективно, приказане су релативне грешке модела за процену транзиционе активности. Бинарни делитељи су различитих величина, и статистика (аутокорелација) улазних сигнала дељеника и делиоца. Из табела се може видети да су све релативне грешке мање од 10% и да је модел у стању да даје тачне вредности за процену транзиционе активности бинарних делитеља.

Треба рећи и то да се *fractional* остатак добија тако што се целобројни остатак поново подели делиоцем и да се хардверска структура бинарног делитеља који даје остатак *fractional* типа уствари састоји од две хардверске структуре бинарног делитеља са целобројним остатком (види слику 4.18). Улази у једну хардверску структуру су дељеник и делилац док се на улаз другог кола за бинарно дељење воде целобројни остатак првог дељења и делилац.



Слика 4.18: Бинарни делитељ са *fractional* остатком: (а) симбол; (б) блок шема

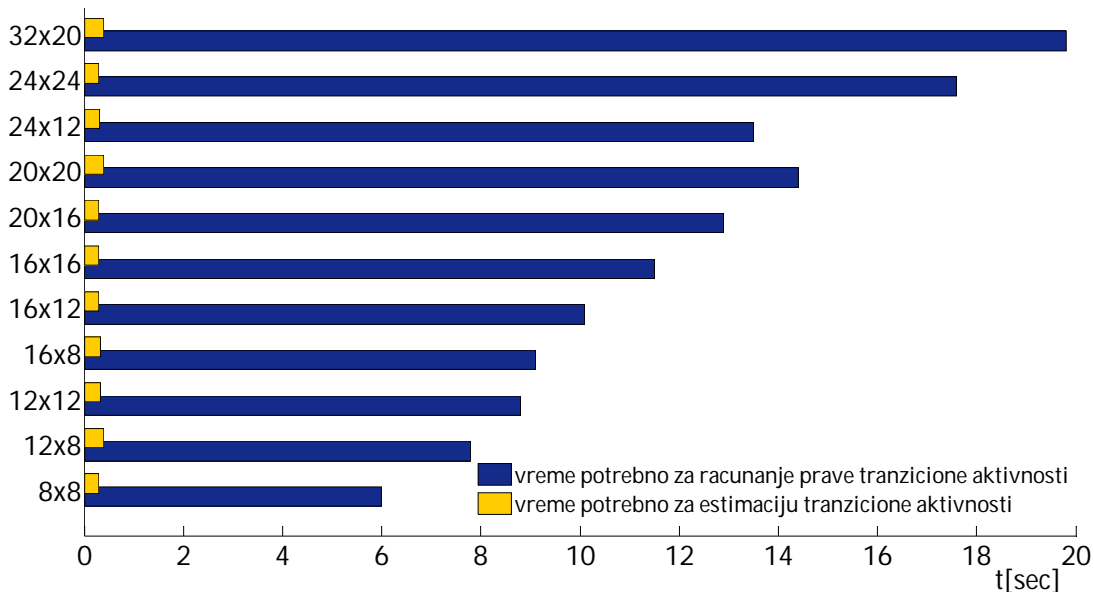
Даље, са слика 4.19 и 4.20 на којима су, за бинарне делитеље са целобројним и *fractional* остатком, приказана просечна времена рачунања транзиционе активности може се видети да је наш модел много бржи и да треба време реда милисекунде да би проценио транзициону активност бинарног делитеља. С друге стране, за добијање праве вредности транзиционе активности потребно је време реда величине од неколико секунди до неколико десетина секунди. Из табела се такође види да је за рачунање праве вредности транзиционе активности *fractional* делитеља потребно дупло више времена у односу на делитељ са целобројним остатком. Ови резултати су очекивани, имајући у виду да хардверску структуру *fractional* делитеља чине два делитеља са целобројним остатком.

Табела 4.4: Релативна грешка (изражена у %) процењене у односу на стварну транзициону активност делитеља са целобројним остатком

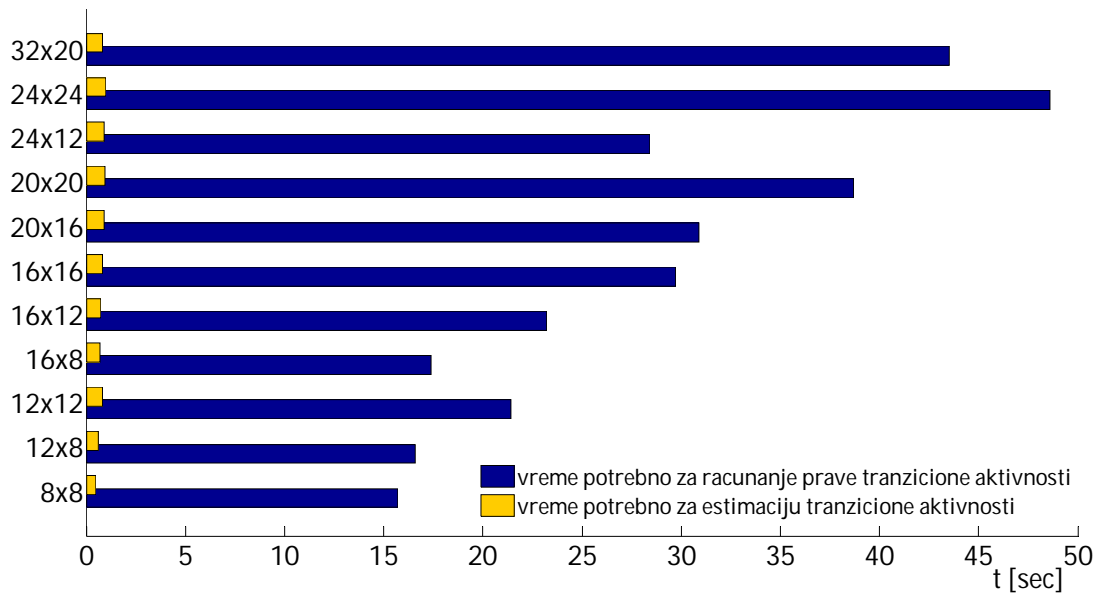
size \ ρ	0	0.5	0.9	0.95	0.99	0.9995
8x8	-9.67	-9.62	-8.34	-3.05	1.38	5.86
12x8	-8.17	-8.20	-5.38	-1.60	2.56	1.62
12x12	-9.82	-9.29	-7.45	-3.21	0.43	1.59
16x8	-6.28	-5.87	-3.67	-0.57	2.46	1.15
16x12	-8.35	-7.4	-4.94	-1.66	1.18	1.83
16x16	-8.39	-7.9	-5.83	-2.74	0.42	0.59
20x16	-7.68	-7.42	-4.98	-2.44	0.36	0.86
20x20	-7.72	-7.45	-5.45	-2.85	-0.02	0.53
24x12	-6.11	-5.34	-3.53	-1.38	1.10	1.89
24x24	-6.67	-6.37	-4.52	-2.39	0.09	0.50
32x20	-5.38	-5.02	-3.24	-1.59	0.35	0.68

Табела 4.5: Релативна грешка (изражена у %) процењене у односу на стварну транзициону активност делитеља са *fractional* остатком

size \ ρ	0	0.9	0.99	0.9995
16x16	8.9	5.8	1.7	1.1
12x12	10.7	6.9	1.8	-4.4
8x8	10.2	5.7	-4.5	-6.3
16x12	10.1	6.5	2.8	-1.2
16x8	8.7	4.8	-2.9	3.4
12x8	9.5	5.6	-3.1	3.2
20x16	8.5	5.5	2.3	2.7
20x20	7.7	4.9	1.8	1.0
24x12	8.7	5.8	2.8	-1.0
24x24	6.7	4.3	1.5	1.0
32x20	6.7	4.4	2.2	2.2



Слика 4.19: Просечно време рачунања укупне транзиционе активности у структури бинарног делитеља са целобројним остатком



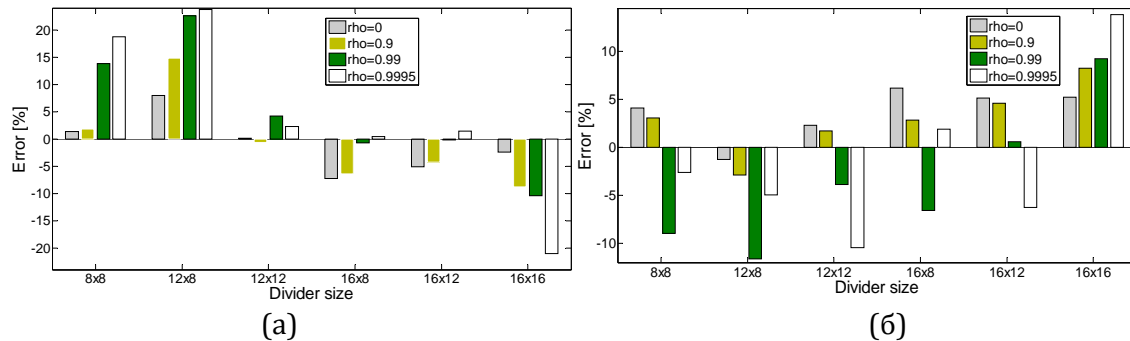
Слика 4.20: Просечно време рачунања укупне транзиционе активности у структури бинарног делитеља са *fractional* остатком

Евалуација модела за процену динамичке потрошње бинарног делитеља

Овај експериментални сет има за циљ да изврши евалуацију модела за процену динамичке потрошње бинарног делитеља представљеног једначином 4.1 и може се поделити на два подсета: ради евалуације модела за процену потрошње, један подсет за референцу користи Virtex II Pro FPGA чип и праве (измерене) вредности потрошње електричне енергије. Референца другог подсета је Virtex 5 FPGA и XPower Analyzer алат за процену потрошње.

Први сет експеримената вршен је над делитељима имплементираним као IP језгра на Xilinx-ову Virtex II Pro XC2VP30-7 FPGA компоненту која се налази на XUP развојној плочи. Аутокорелација улазних сигнала варирала је од 0 до 0.9995 а ширине улаза између 8 и 16 битова. Најпре је мерена потрошња ових делитеља уз помоћ мерне поставке представљене у глави 3. Резултати мерења процесирају се на начин описан у глави 3 како би се измерена вредност потрошње расчланила на динамичку потрошњу логике, интерконеција, дистрибуционе мреже такта и статичку потрошњу. Максималне величине имплементираних делитеља биле су ограничене бројем I/O пинова развојних плоча које су нам у мерној поставци биле на располагању. Након мерења и процесирања резултата, измерена динамичка потрошња логике поредила се са процењеном вредношћу за динамичку потрошњу. Резултати, за делитеље са целобројним и *fractional* остатком, приказани су на слици 4.21. Процењена вредност добија се кад се, према једначини 4.1, вредност процењене укупне транзиционе активности помножи са емпиријски добијеном константом a . За одређивање емпиријске константе a , у случају делитеља са целобројним остатком, коришћено је 9 паровао (P_i , SW_i). Величина бинарних делитеља коришћених за одређивање ове константе била је 8x8, 12x12 и 16x16. За сваки делитељ аутокорелација улазних сигнала била је 0, 0.9, и 0.9995 (скуп ових делитеља осенчан је и уоквирен испрекиданом линијом на

слици 4.21a). За константу a емпиријски је утврђено да је једнака $220.21\mu W$. За бинарни делитељ са остатком типа *fractional* за константу a добијено је $133,58\mu W$. Мања вредност константе a код *fractional* делитеља последица је њихове веће транзиционе активности у односу на делитеље са целобројним остатком. Повећана транзициона активност потиче од њихове веће хардверске сложености.



Слика 4.21: Релативне грешке модела за процену динамичке потрошње у односу на измерену потрошњу (а) за бинарни делитељ са целобројним остатком, (б) за бинарни делитељ са *fractional* остатком

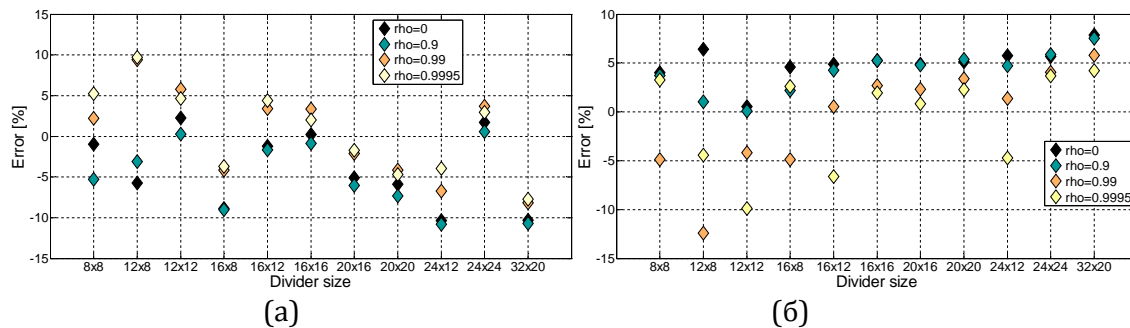
Са слике 4.21 може се видети да је методологија процене потрошње прилично тачна будући да највећи број грешака лежи испод 15%, са средњом релативном грешком од 7.5% за случај делитеља са целобројном константом и 5.35% за бинарни делитељ са остатком типа *fractional*. Ове мале грешке потврда су исправности претпоставке да у структури бинарног делитеља нема много гличева и да их је могуће занемарити приликом процене потрошње електричне енергије.

Други сет експеримената вршен је над делитељима имплементираним као IP језгра на Xilinx-ову Virtex 5 XC5VLX30 FPGA компоненту. Услед недостатка мерне опреме (развојне плоче) која садржи FPGA компоненту из ове фамилије, за добијање референтне вредности динамичке потрошње користили смо XPower Analyzer из ISE 11.1 верзије софтвера. Најпре је, уз помоћ ModelSim програмског пакета, извршена симулација дизајна, и то за по 10000 вредности на улазима дељеника и делиоца. Након симулације креира се *.saif* фајл који садржи информације о транзиционим активностима свих чворова и елемената у дизајну. Заједно са *.ncd* и *.pcf* фајловима који садрже информације и размештеност и повезаност дизајну *.saif* фајл се прослеђује XPower Analyzer алату. Пошто XPower алат генерише детаљан извештај о потрошњи (види поглавље 3.2.3) у MATLAB програмском пакету врши се обрада *.txt* фајла XPower извештаја како би се од укупне динамичке потрошње дизајна добила динамичка потрошња логике (динамичка потрошња глобалних интерконекција се не узима у обзир).

Будући да смо променили FPGA компоненту на коју имплементирамо дизајн бинарног делитеља потребно је поново извршити калибрацију модела тј. емпиријски одредити константу a . Ова константа одређена је на исти начин и са истим делитељима и коефицијентима аутокорељације као и у случају Virtex II Pro компоненте. За константу a , за делитељ са целобројним остатком, у случају Virtex 5 XC5VLX30 компоненте добијено је $a=40\mu W$. Мања константа a доказ је мање динамичке потрошње Virtex 5 FPGA чипа у односу на старију Virtex II Pro фамилију. Тренд смањивања динамичке потрошње сваке нове генерације FPGA

компоненте у односу на компоненте из претходних генерација (услед смањивања напона напајања, димензија транзистора и капацитивности чворова) илустрован је на слици 2.17. Емпиријска константа a код *fractional* делитеља једнака је $27,33\mu W$.

Пошто су нам на располагању референтне вредности динамичке потрошње логике бинарног делитеља (добили смо уз помоћ XPower Analyzer алата) извршили смо њихово поређење са вредностима динамичке потрошње добијеним применом модела за процену динамичке потрошње бинарног делитеља. Релативна грешка процењене вредности динамичке потрошње у односу на референтну вредност (за делитеље са целибројним и *fractional* остатком) приказана је на слици 4.22.



Слика 4.22: Релативне грешке модела за процену динамичке потрошње у односу на референтну XPower потрошњу (а) за бинарни делитељ са целибројним остатком, (б) за бинарни делитељ са *fractional* остатком

Може се видети да је наш модел за процену динамичке потрошње у стању да обезбеди врло тачне процене динамичке потрошње електричне енергије. За случај бинарног делитеља са целибројним остатком максимална и средња релативна грешка једнаке су 10.8% и 4.82%, респективно, док су за случај делитеља са остатком *fractional* типа ове грешке 12.43% и 4.34%.

У табели 4.6, за бинарне делитеље са целибројним и *fractional* остатком, приказана су упоредна времена потребна за добијање процењене вредности динамичке потрошње. У односу на наш модел за процену потрошње, XPower алат (заједно са ModelSim симулатором и ISE алатом за имплементацију) је значајно спорији.

Табела 4.6: Упоредна времена потребна за процену динамичке потрошње електричне енергије бинарних делитеља са целибројним и *fractional* остатком

	Делитељ са целибројним остатком			Делитељ са <i>fractional</i> остатком		
	8x8	16x16	32x20	8x8	20x16	24x24
ISE+XPower+ModelSim	2m9s	7m34s	18m19s	3m7s	9m50s	18m5s
a*SWest	0.12s	0.13s	0.12s	0.14s	0.15s	0.15s

Глава 5

Закључци

ИСТРАЖИВАЧКИ рад представљен у овој дисертацији грубо се може поделити у два дела. Први део дисертације (главе 1 и 2) бави се проблемима оптимизације тј. минимизације (како статичке тако и динамичке) потрошње електричне енергије, док је други део (главе 3 и 4) посвећен моделима за процену динамичке потрошње електричне енергије DSP (Digital Signal Processing) дизајна.

5.1. Минимизација потрошње електричне енергије: закључци

Много је разлога због којих дизајнери и развојни инжењери брину о дисипацији енергије данашњих интегрисаних кола. Диктирана захтевима за повећањем функционалности и перформанси, и у исто време смањивањем цене, минијатуризација тј. скалирање силицијумске технологије одвија се преко четрдесет година. У почетку је смањивање димензија транзистора имало само позитивне ефекте: смањивала се површина силицијума потребна за имплементацију неког интегрисаног кола, повећавала се његова брзина рада уз истовремено смањивање потрошње. Кад је број транзистора по јединици површине достигао критичну тачку почели су да се јављају до тада непознати ефекти који су повећавали потрошњу (нарочито статичку) електричне енергије и утицали на то да минимизација потрошње, поред брзине и површине, постане још један критеријум за дизајнере савремених VLSI кола.

Смањивање димензија CMOS транзистора за последицу има следеће ефекте кратког канала (Short Channel Effects - SCE):

- Засићење брзине носилаца тј. линеарна зависност струје дрејна и напона између гејта и сорса. Последица овога је смањена струјна моћ транзистора.

- Напон прага транзистора V_{th} функција је дужине канала L и радних напона, што за последицу има смањену могућност контроле напона прага преко поларисања субстрата.
- Повећане струје цурења (како струја цурења у каналу тако и струја тунеловања гејта) утичу на то да се однос активне и цуреће струје (I_{on}/I_{off}) смањује. Овај однос се смањује како због смањивања активне тако и због повећања цуреће струје.

У лабораторијама широм света развијају се нове транзисторске структуре које имају за циљ да одговоре на изазове повећане потрошње електричне енергије. Те нове структуре углавном имају неко од следећих својстава: повећану покретљивост носилаца наелектрисања, бољу контролу напона прага транзистора или смањене струје цурења. Треба такође рећи и то да су и ове нове транзисторске структуре ограниченог трајања тј. да ће омогућити само привремено (за генерацију или две) побољшање технолошког процеса производње.

Проблем оптимизације (минимизације) потрошње је вишедимензионалан, што значи да се до дизајна са оптималном потрошњом електричне енергије може доћи једино онда када се о потрошњи води рачуна у свакој фази дизајнирања система. Такође, оптимизационе технике морају међусобно бити избалансиране како би минимизациони потенцијали сваког од параметара који утиче на потрошњу били максимално искоришћени. Треба имати на уму и то да оптимално не значи увек и минимално зато што су минимална потрошња електричне енергије и минимално кашњење дизајна међусобно опречни захтеви. Тако се може десити да обраћањем пажње само на минимизацију потрошње добијемо дизајн са неприхватљивим кашњењем. Потребно је дакле правити компромисе, па коректна оптимизација типично или минимизира потрошњу енергије за унапред задата ограничења у погледу брзине рада или максимизира процесну моћ за унапред задату дозвољену потрошњу.

Из свега наведеног може се закључити да смо закорачили у еру у којој је, за разлику од ранијих времена када су дизајнери обраћали пажњу само на перформансе система, потрошња интегрисаних кола постала ограничавајући фактор њиховог даљег развоја. Такође, мало је вероватно да ће у скорој будућности (бар не до 2030. године [Nas11]) CMOS структуру заменити нека нова транзисторска структура која ће решити текуће проблеме (слично преласку са биполарних на CMOS транзисторе). Стога је неопходно развијати нове стратегије дизајнирања дигиталних интегрисних кола и побољшавати постојеће архитектуре за обраду података.

5.2. Процена динамичке потрошње електричне енергије: закључци

У дизајнима који се састоје од стазе података и управљачке јединице потрошња логичких компонената које обављају аритметичке операције доминира у односу на потрошњу других компоненти у дизајну. Део ове дисертације који се бави моделима за процену динамичке потрошње наставак је рада описаног у [Jev09], у коме су представљени аналитички модели високог нивоа за процену динамичке

потрошње логике бинарних сабирача и множача (*embedded* типа и имплементираних у LUT областима FPGA структуре).

У овој дисертацији представљено је побољшање тачности модела за процену динамичке потрошње логике бинарног множача (LUT имплементације *ripple carry array* архитектуре). Побољшање је остварено тиме што је, за разлику од претходног модела, у обзир узета нелинеарност у транзиционој активности на нивоу бита која се испољава на излазу множача. Експерименталним путем утврђено је да модел предложен у дисертацији даје значајно тачније (у неким случајевима 4-5 пута) процене динамичке потрошње.

Поред побољшања тачности модела множача у дисертацији су представљени и модели за процену транзиционе активности (тј. динамичке потрошње) различитих типова бинарних делитеља *Radix-2 non-restoring* архитектуре.

Главни закључци могу се сумирати на следећи начин:

- Услов за примену аналитичких модела за процену динамичке потрошње електричне енергије је да су улазни сигнали операнада стационарни и да имају Гаусову расподелу (тј. да су одређени средњом вредношћу, варијансом и коефицијентом аутокорељације). Ова претпоставка може се сматрати исправном будући да је Gauss-ова расподела често коришћена за тачно моделовање DSP сигнала [LR95, JKS+99, CGC05].
- Аналитички модели за процену транзиционе активности унутар чворова логичке структуре су параметризовани. На основу броја битова улазних сигнала и њихове статистике (средње вредности, варијансе и аутокорељације) извршена је, према транзиционој активности њихових битова, подела улазних сигнала на регионе.
- За рачунање укупне транзиционе активности логичке структуре чија се процена потрошње врши развијен је пробабилистички метод. Применом овог метода рачунају се транзиционе активности и вероватноће да су у стањима '0' или '1' излаза основних ћелија логичке структуре. Укупна транзициона активност логичке структуре једанак је збиру транзиционих активности свих ћелија.
- У случају бинарних делитеља, примена пробабилистичке методологије за процену транзиционе активности даје прецењене (много веће у односу на стварне) вредности. Узрок овога је специфичност логичке структуре бинарног делитеља тј. међусобна корелација која се јавља на улазима неких логичких компоненти у његовој структури. Због тога је извршена мала модификација методологије за прорачун транзиционе активности која је дала задовољавајуће резултате.
- Главне предности презентоване методологије за процену динамичке потрошње су њена брзина (процене динамичке потрошње добијају се за време реда стотинак милисекунди) и лакоћа калибрације (за добијање емпиријске константе a потребно је обавити десетак мерења потрошње).
- Поред брзине и лакоће калибрације методологија се одликује и великом тачношћу. У поређењу са стварним вредностима измерене динамичке потрошње средње релативне грешке (како модела бинарног множача тако и делитеља) не прелазе 10%.

- Будући да су им за процену динамичке потрошње потребне само информације о броју битова улазних сигнала и њиховим статистикама, представљене моделе високог нивоа могуће је успешно користити у алгоритмима високог нивоа који имају за циљ оптимизацију потрошње електричне енергије.

5.3. Будући рад

Рад представљен у овој дисертацији који се тиче процене динамичке потрошње у обзир узима само модуле са регистрованим улазима и излазима. За проширење применљивости методологије на архитектура са нерегистрованим модулима потребно је проширити постојеће моделе и развити методологију за моделовање гличева зато што ће у архитектурама оваквог типа добар део укупне транзиционе активности да буде последица гличева.

Да би модел за процену потрошње био комплетан неопходно је развити и методологије за процену потрошње интерконекција као и статичке потрошње дизајна будући да статичка потрошња полако постаје доминантија у односу на динамичку потрошњу.

Један од циљева који се тиче процене потрошње интерконекција је развој ефикасног *floorplanning* алгоритма који би омогућио успешну интеграцију модела за процену потрошње интерконекција у оптимизационе технике високог нивоа, заједно са предложеним моделима за процену динамичке потрошње логике који су већ применљиви на високом нивоима апстракције система.

Поред логичких компоненти описаних у [Јев09] и у овој дисертацији, у пракси се врло често наилази и на компоненте које обављају аритметичке операције над операндима од којих је један константан (приликом имплементације дигиталних FIR филтара, у дигиталним телекомуникацијама итд.). Примена описане пробабилистичке методологије у овим случајевима не би дала задовољавајуће резултате будући да један од сигнала нема Гаусову расподелу. Због тога би будући рад на пољу методологије процене потрошње електричне енергије могао да укључи и прилагођење постојећих логичких модела структурама које обављају аритметичке операције над операндима од којих је један константан.

Додаци

Додатак А

А.1. Вредност отпорности за мерни систем

Систем за мерење потрошње састоји се из две плоче: Xilinx-ове XUP и Altera-ине DSP (мерни систем приказан је на слици 1). Мерење потрошње језгра XUP плоче остварено је мерењем напона на отпорности која је смештена иза извора за напајање језгра FPGA чипа. Вредност отпорности изабрана је тако да не наруши исправност функционисања регулатора напона напајања на XUP плочи. У наставку је дато објашњење како се долази до вредности отпорности.

1.5V за напајање FPGA језгра добијено је уз помоћ синхроног *buck-switching* регулатора који се екстерно напаја са 4.5V–5.5V [Xil05]. Стабилна и фиксна вредност на излазу регулатора добијена је уз помоћ петље повратне спреге, тако што се са пина за напајања језгра XC2VP30 FPGA компоненте (тачка В на слици 2а) повратна спрега води на улаз регулатора. Линија повратне спреге на слици 2а је подељана и испрекидана. Напон за напајање FPGA језгра на улазном пину FPGA чипа, V_B , одржава се на 1.5V и тиме се обезбеђује исправна функционалност чипа.

Упрошћена блок шема *buck-switching* регулатора са импулсно-ширинском (PWM) модулацијом приказана је на слици 2б. Пошто се повратна спрега узима из тачке В на слици 2а, вредност напона на излазу регулатора (означена са А на слици 2) биће:

$$V_A = V_B + V_R = V_B + R \cdot i \quad (1)$$

Будући да је регулатор *buck-switching* типа битно је да се избегне засићење унутрашњег калема регулатора. Ово засићење се дешава онда када просечна вредност напона на излазу *buck-switching* регулатора превазилази просечну вредност напона на другом крају калема [ED01] (тачка D на слици 2б). Просечна вредност напона у тачки D једнака је:

$$V_D = d \cdot V_{in} \quad (2)$$

где је d фактор испуне (duty cycle) *buck-switching* регулатора. Према томе, засићења калема неће бити све дотле док је напон у тачки А, V_A , мањи или једнак максималном напону V_D . Из једначина 1 и 2 добијамо услов који је потребно испунити:

$$V_B + R \cdot i \leq d_{\max} \cdot V_{in} \quad (3)$$



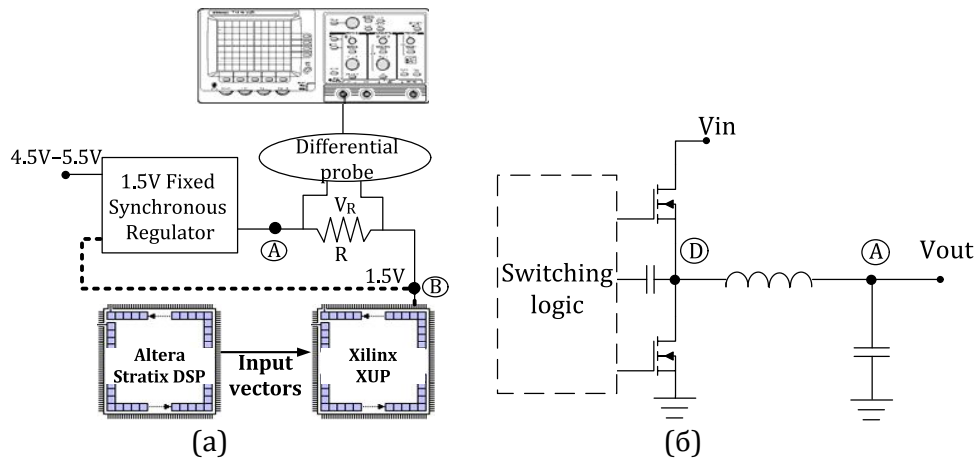
Слика 1: Мерни систем

Будући да вредност d_{\max} није дата од стране произвођача регулатора, експериментално је добијено да је једнака 0.5. Пошто је номинална вредност улазног напона V_{in} једнака 5V, а вредност напона у тачки А представља збир напона на отпорнику и напона у тачки В (1.5V), добијамо да просечна вредност напона на отпорнику, $V_R = V_A - V_B$, не сме прећи вредност од 1V. Имајући у виду и израз за потрошњу FPGA чипа:

$$P = V_B \cdot i_B = 1.5 \cdot \frac{V_R}{R} \quad (4)$$

закључујемо да је потрошња чипа обрнуто сразмерна отпорности. Што се већа потрошња дизајна очекује, то мању отпорност R треба ставити. У принципу, за одређивање вредности отпорности треба знати очекивану вредност потрошње имплементираних дизајна. Имајући у виду тај податак, и чињеницу да напон на отпорнику не сме да буде већи од 1V, лако је прорачунати минималну вредност за отпорност R која обезбеђује исправно функционисање мерног система. За мерење потрошње нашег DSP3 дизајна очекивана вредност потрошње добијена је коришћењем XPower Analyzer CAD алата. На основу тога, за минималну вредност отпорности добијено је $R_{\min} = 6.4\Omega$, па је за отпорност узета прва следећа већа вредност из E24 отпорног низа. Отпорници у овом низу су са толеранцијом од $\pm 5\%$.

Треба рећи и то да је мерни систем власништво Политехничког Универзитета у Мадриду где је и осмишљен за потребе њиховог истраживања.



Слика 2: (а) Мерна поставка (б) Buck-switching PWM регулатор

A.2. MARWEL

MARWEL је софтверски алат писан у C++ програмском језику и развијен на Политехничком Универзитету у Мадриду. Користи се за екстракцију информација о рутирању дизајна имплементираних на Xilinx FPGA компоненте. Фајл над којим MARWEL алат обавља анализу је *.xdl* формата и представља текстуалну верзију размештеног и рутираног дизајна. Креира се од стране Xilinx XDL алата.

A.2.1. Структура XDL фајла

.xdl фајл добија се након следеће Xilinx ISE команде:

```
-ncd2xdl SourceFile.ncd OutputFile.xdl (5)
```

Треба рећи и то да је инверзном командом (*-xdl2ncd*) могуће конвертовати *.xdl* текстуални фајл у *.ncd* графичку репрезентацију размештеног и повезаног дизајна.

.xdl фајл састоји се из два дела. У првом делу налази се листа свих инстанци дизајна, заједно са њиховом конфигурационим подацима и локацији унутар FPGA. Инстанца дизајна припада једној од следећих група: логички блокови, I/O пинови, DCM-ови (Digital Clock Manager) и мултиплексери. Опис сваке инстанце почиње речју "inst" (види слику 3). Након тога иде име инстанце које се касније користи за опис свих веза са којима је ова инстанца повезана преко било ког од својих пинова. Име инстанце праћено је информацијама о типу инстанце и позицију унутар FPGA. Ако је инстанца смештена унутар CLB блока онда се специфицира и слајс резервисан за инстанцу.

Основни опис инстанце праћен је њеним конфигурационим детаљима који нису релевантни за екстракцију информација о рутирању дизајна (MARWEL алат користи само прву линију описа сваке инстанце, уоквирена на слици 3).

Други део XDL фајла садржи листу свих веза у дизајну. Један пример описа везе дат је на слици 4. Опис увек почиње речју "net" праћеном именом везе. Након тога даје се име пина од кога веза почиње, као и листа пинова на којима се веза

завршава. Имена пинова су уствари имена неких инстанци чији се опис налази у првом делу *.xdl* фајла. Важно је напоменути да, за сваку везу, постоји само један “излазни” пин (почетак везе) док “улазних” пинова (крајева везе) може бити и више.

Instance name	Logic or I/O	Instance CLB position	Instance slice position
inst "m/N1479"	"SLICE"	placed R42C41	SLICE_X81Y77

```

cfg " BXINV::#OFF BXOUTUSED::#OFF BYINV::#OFF
BYINVOUTUSED::#OFF BYOUTUSED::#OFF
CEINV::#OFF CLKINV::#OFF COUTUSED::0
CY0F::PROD CY0G::PROD CYINIT::CIN CYSELF::F
CYSELG::G DIF_MUX::#OFF DIGUSED::#OFF
DIG_MUX::#OFF DXMUX::#OFF
DYMUX...

```

Слика 3: Синтакса XDL фајла: први део

	Line number
Net name	1
Net end	2
Net beginning	3
Switch matrix	4
CLB position	5
	6
	7
	8
	9
	10
	11

```

net "operand1<1>" , Wire
inpin "m/N214" BX , description
outpin "input1<1>" IQ1 ,
pip RIOIR38 I_Q10 -> OMUX5 ,
pip R39C46 OMUX_SW5 -> S6BEG9 ,
pip R45C46 S6END9 -> W2BEG8 ,
pip BRAMR44C8 W2END_N8 -> W2BEG0 ,
pip R44C43 W2END0 -> W2BEG0 ,
pip R44C41 W2END0 -> W2BEG0 ,
pip R44C40 W2MID0 -> BX0 ,
pip R44C40 BX0 -> BX_PINWIRE0 ,
;

```

Слика 4: Синтакса XDL фајла: други део

Идентификатор “pip” користи се да опише везу унутар матрице програмабилних прекидача (switch matrix). Праћен је информацијама о позицији програмабилне матрице (са истом нотацијом која се користи и приликом позиционирања CLB блока у првом делу *.xdl* фајла). На крају је дат опис електричних веза које су повезана унутар једне матрице прекидача. Позиција матрице прекидача као и описи веза врло су значајне информације приликом екстракције информација о повезивању.

Позиција CLB блока

Као што је већ напоменуто, FPGA компонента представља дводимензионални низ конфигурабилних логичких блокова (CLB). Позиција сваког CLB блока одређена је његовим координатама тј. информацијама о броју врсте и колоне у дводимензионалном низу. На пример, опис CLB блока уоквиреног на слици 4 почиње словом *R* и бројем који представља координату врсте. Исто важи и за слово *C* и за координату колоне. За CLB блокове који имају специјалне локације, осим информација о координатама користи се још неколико нотација. На пример, у четвртој линији на слици 4 локација CLB блока је *RIOIR38*. Ово је нотација блока који садржи I/O пинове и који је смештен на десној страни чипа (прво *R* значи да је блок на десној страни а друго врсту у FPGA низу). Како је број колоне блока који је смештен уз десну страну FPGA структуре уствари максимална вредност координате колоне, једина потребна информација за дефинисање позиције блока је позиција (координата) врсте. Поред ове класе CLB блокова постоји још неколико других типова блокова: CLB блокови смештени у угловима FPGA

структуре, CLB блокови смештени поред уграђених (embedded) блокова, CLB блокови са матрицама прекидача итд.

Опис електричне везе

Четири типа глобалних веза (линија) за повезивање описане су на следећи начин:

- Директне (*single*) линија: почиње нотацијом *OMUX* након које следи број за праћење и/или правац, у зависности од тога да ли се означава почетак или крај линије. На пример, 4. ред на слици 4 означава почетак *single* линије са бројем 5, док се у 5. реду (на слици 4) може видети да се ова линија завршава правцем југозапад (*SW* – south west).
- *Double* линија: почиње информацијом о правцу праћеном бројем 2 (од *double*) и скраћеницом *BEG*, *MID* или *END* (да означи почетњи, средњи или задњи крај линије, респективно). На крају нотације је број за праћење (видети, на пример, 6. ред на слици 4 где *W2BEG8* означава *double* линије са бројем за праћење 8 и правцем према западу).
- *Hex* линија: исте је структуре као и *double* линија. На пример, заокружена нотација у 6. реду на слици 4 (*S6END9*) означава крај *hex* линије са бројем за праћење 9 и правцем према југу.
- *Long* линија: постоје две нотације за *long* линију: *LV* или *LH*, у зависности од тога да ли је њихов правац хоризонталан *H* или вертикалан *V*.

Додатне информације о структури *MARWEL*-а заинтересовани читалац може наћи у [Jev09].

А.3. Примена пробабилистичког метода на ћелије сабирача, инвертора и мултиплексера

За рачунање транзиционе активности структура бинарних сабирача, инвертора и мултиплексера коришћен је пробабилистички метод. Овај метод третира транзициону активност као вероватноћу да ће сигнал променити своје логичко стање за време једног тактног интервала. Поред транзиционе активности, пробабилистички метод барата и са вероватноћама да сигнал има вредност логичке '0' или логичке '1'. Сваки логички гејт представљен је једначином која описује зависност транзиционе активности и вероватноће излазног бита од транзиционих активности и вероватноћа улазних бита гејта. Уз помоћ табела истинитости (приказаних на слици 5), на начин који је за ћелију потпуног бинарног сабирача објашњен у поглављу 4.5.2, добијени су аналитички изрази за транзиционе активности и вероватноће да су у стањима '0' и '1' излазних и *carry* бита.

За ћелију полусабирача (НА) приказану на слици 5 аналитички изрази за транзиционе активности и вероватноће излазног и *carry* бита су следећи:

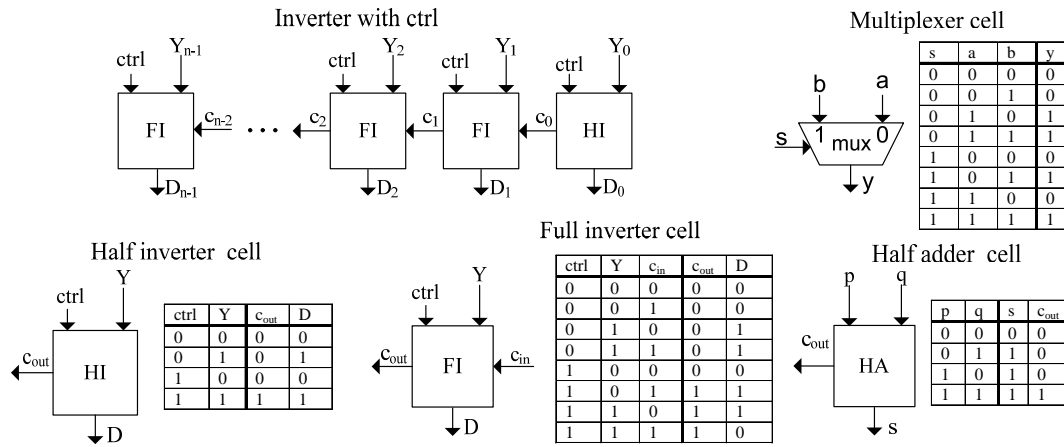
$$\begin{aligned} s &= p^0 \cdot q^0 \cdot (p - p \cdot q + q) + p^0 \cdot q^1 \cdot q \cdot (1 - p) + p^1 \cdot q^0 \cdot p \cdot (1 - q) + p^1 \cdot q^1 \cdot p \cdot q \\ c_{out} &= p^0 \cdot q^0 \cdot p \cdot q + p^0 \cdot q^1 \cdot p \cdot (1 - q) + p^1 \cdot q^0 \cdot q \cdot (1 - p) + p^1 \cdot q^1 \cdot (p + q - p \cdot q) \end{aligned} \quad (6)$$

$$\begin{aligned}
s^1 &= 1 - p^0 \cdot q^0 \\
s^0 &= 1 - s^1 \\
c^0 &= 1 - p^1 \cdot q^1 \\
c^1 &= 1 - c^0
\end{aligned}
\tag{7}$$

У случају логичке структуре инвертора са контролним сигналом коришћена је следећа листа појмова:

- $ctrl, Y, C_{in}, C_{out}$ и D су вероватноће транзиција свих улазних и излазних прикључака ћелије инвертора.

- $ctrl^0, ctrl^1, Y^0, Y^1, C_{in}^0, C_{in}^1, C_{out}^0, C_{out}^1, D^0, D^1$ су вероватноће улазних и излазних сигнала да су у стањима логичке '0' логичке '1', респективно.



Слика 5: Логичке структуре и табеле истинитости ћелија инвертора и мултиплексера

Конечан израз за транзициону активности ћелије потпуног инвертора (FI) има следећи облик:

$$c_{out} = c_{out}^{A1} + c_{out}^{A2} + c_{out}^{B1} + c_{out}^{B2} + c_{out}^{B3} + c_{out}^C \tag{8}$$

$$c_{out}^{A1} = [ctrl^0 \cdot Y^0 \cdot c_{in}^1 + ctrl^0 \cdot Y^1 \cdot c_{in}^0 + ctrl^0 \cdot Y^1 \cdot c_{in}^1 + ctrl^1 \cdot Y^1 \cdot c_{in}^1] \cdot ctrl \cdot (1 - Y) \cdot (1 - c_{in})$$

$$c_{out}^{A2} = ctrl^1 \cdot Y^0 \cdot c_{in}^0 \cdot [(1 - ctrl) \cdot (1 - Y) \cdot c_{in} + (1 - ctrl) \cdot Y \cdot (1 - c_{in})] + ctrl^1 \cdot Y^0 \cdot c_{in}^1 \cdot [(1 - ctrl) \cdot (1 - Y) \cdot c_{in} + ctrl \cdot (1 - Y) \cdot (1 - c_{in})] + ctrl^1 \cdot Y^1 \cdot c_{in}^0 \cdot [(1 - ctrl) \cdot Y \cdot (1 - c_{in}) + ctrl \cdot (1 - Y) \cdot (1 - c_{in})]$$

$$c_{out}^{B1} = ctrl^1 \cdot Y^1 \cdot c_{in}^1 \cdot [ctrl \cdot Y \cdot (1 - c_{in}) + ctrl \cdot (1 - Y) \cdot c_{in} + (1 - ctrl) \cdot Y \cdot c_{in}]$$

$$c_{out}^{B2} = ctrl^0 \cdot Y^0 \cdot c_{in}^1 \cdot ctrl \cdot Y \cdot (1 - c_{in}) + ctrl^0 \cdot Y^1 \cdot c_{in}^0 \cdot ctrl \cdot (1 - Y) \cdot c_{in} + ctrl^1 \cdot Y^0 \cdot c_{in}^0 \cdot (1 - ctrl) \cdot Y \cdot c_{in}$$

$$c_{out}^{B3} = [ctrl^0 \cdot Y^0 \cdot c_{in}^0 + ctrl^0 \cdot Y^1 \cdot c_{in}^1 + ctrl^1 \cdot Y^0 \cdot c_{in}^1 + ctrl^1 \cdot Y^1 \cdot c_{in}^0] \cdot [ctrl \cdot Y \cdot (1 - c_{in}) + ctrl \cdot (1 - Y) \cdot c_{in}]$$

$$c_{out}^C = (1 - ctrl^0 \cdot Y^1 \cdot c_{in}^1 + ctrl^1 \cdot Y^0 \cdot c_{in}^0) \cdot ctrl \cdot Y \cdot c_{in}$$

$$\begin{aligned}
D &= (ctrl^0 \cdot Y^0 \cdot c_{in}^0 + ctrl^0 \cdot Y^1 \cdot c_{in}^0) \cdot (Y + ctrl \cdot c_{in} - 2 \cdot ctrl \cdot Y \cdot c_{in}) + ctrl^0 \cdot Y^0 \cdot c_{in}^1 \cdot \\
&\cdot (ctrl + Y - ctrl \cdot Y - ctrl \cdot c_{in} - 2 \cdot ctrl \cdot Y \cdot c_{in}) + ctrl^0 \cdot Y^1 \cdot c_{in}^1 \cdot (ctrl + Y - 2 \cdot ctrl \cdot Y - \\
&- ctrl \cdot c_{in} + 2 \cdot ctrl \cdot Y \cdot c_{in}) + ctrl^1 \cdot Y^0 \cdot c_{in}^0 \cdot (Y + c_{in} - 2 \cdot Y \cdot c_{in} - ctrl \cdot c_{in} + 2 \cdot ctrl \cdot Y \cdot c_{in}) + \\
&+ (ctrl^1 \cdot Y^0 \cdot c_{in}^1 + ctrl^1 \cdot Y^1 \cdot c_{in}^1) \cdot (ctrl + Y + c_{in} - 2 \cdot ctrl \cdot Y - 2 \cdot Y \cdot c_{in} - ctrl \cdot c_{in} + \\
&+ 2 \cdot ctrl \cdot Y \cdot c_{in}) + ctrl^1 \cdot Y^1 \cdot c_{in}^0 \cdot (Y + c_{in} - 2 \cdot Y \cdot c_{in} - ctrl \cdot c_{in} + 2 \cdot ctrl \cdot Y \cdot c_{in})
\end{aligned}$$

$$\begin{aligned}
c_{out}^1 &= ctrl^1 \cdot (1 - Y^0 \cdot c_{in}^0) \\
c_{out}^0 &= (1 - c_{out}^1) \\
D^1 &= ctrl^0 \cdot Y^1 + ctrl^1 \cdot (Y^1 c_{in}^0 + Y^0 c_{in}^1) \\
D^0 &= 1 - D^1
\end{aligned} \tag{9}$$

Што се тиче ћелије полуинвертора (НІ), аналитички изрази за транзициону активност и вероватноће излазног и *carry* бита су следећи:

$$\begin{aligned}
c_{out} &= ctrl^0 \cdot Y^1 \cdot ctrl \cdot (1 - Y) + ctrl^1 \cdot Y^0 \cdot Y \cdot (1 - ctrl) + ctrl^1 \cdot Y^1 \cdot (ctrl + Y - 2 \cdot ctrl \cdot Y) + \\
&+ (ctrl^0 \cdot Y^0 + ctrl^1 \cdot Y^1) \cdot ctrl \cdot Y \\
D &= Y
\end{aligned} \tag{10}$$

$$\begin{aligned}
c_{out}^1 &= ctrl^1 \cdot Y^1 \\
c_{out}^0 &= 1 - c_{out}^1 \\
D^1 &= Y^1 \\
D^0 &= 1 - D^1
\end{aligned} \tag{11}$$

Као што се може видети са слике 5, логичка структура ћелије мултиплексера је простија с обзиром на то да нема пропагације сигнала излазног преноса c_{out} . Аналитички изрази за транзициону активност и вероватноће излаза Y су следећи:

$$\begin{aligned}
y &= s^0 \cdot a^0 \cdot b^0 \cdot (a - a \cdot s + s \cdot b) + (s^0 \cdot a^0 \cdot b^1 + s^0 \cdot a^1 \cdot b^0) \cdot (a + s - s \cdot a - s \cdot b) + \\
&s^0 \cdot a^1 \cdot b^1 \cdot (a - a \cdot b + a \cdot b \cdot s) + (s^1 \cdot a^0 \cdot b^0 + s^1 \cdot a^1 \cdot b^1) \cdot (b - b \cdot s + a \cdot s) + \\
&(s^1 \cdot a^0 \cdot b^1 + s^1 \cdot a^1 \cdot b^0) \cdot (b + s - s \cdot b - s \cdot a)
\end{aligned} \tag{12}$$

$$\begin{aligned}
y^1 &= s^0 \cdot a^1 + s^1 \cdot b^1 \\
y^0 &= 1 - y^1
\end{aligned} \tag{13}$$

Литература

- [AAN03] E. Acar, R. Arunachalam, S. Nassif, „Predicting short-circuit power from timing models,” *Proc. of ASP-DAC*, pp. 277-282, Jan. 2003.
- [ADS+06] K. Agarwal, H. Deogun, D. Sylvester, K. Nowka, „Power Gating with Multiple Sleep Modes,” *ISQED*, pp. 633-637, 2006.
- [AFG+05] C. Ababei, Y. Feng, B. Goplen, H. Mogal, et al., „Placement and Routing in 3D Integrated Circuits,” *IEEE Design&Test of Computers*, vol. 22, no. 6, pp. 520-531, 2005.
- [AFJS07] N. Abdeli, A. Fouilliant, N. Julien, E. Senn, „High level Power Estimation on FPGA,” *IEEE Symp. on Industrial Electronics*, pp. 925-930, June 2007.
- [Ali06] H. Ali, *Carbon Nanotube Interconnects for IC Chips*, MSc Thesis, Massachusetts Institute of Technology, 2006.
- [AMD+94] M. Alidima, J. Monteiro, S. Devadas, A. Ghosh, M. Papaefthimiou, „Precomputation-based sequential logic optimization for low power,” *IEEE Trans on VLSI*, vol. 2, no. 4, pp. 426-435, Dec. 1994.
- [AMK+05] A. Agarwal, S. Mukhopadhyay, C. Him, A. Raychowdhury, K. Roy, „Leakage Power Analysis and Reduction: Models, Estimation and Tools,” *IEEE Proc. on Computers and Design Techniques*, vol. 153, no. 3, pp. 353-368, May 2005.
- [AN04] J. Anderson, F. Najm, „Power estimation techniques for FPGAs,” *IEEE Trans. on VLSI*, vol. 10, no. 12, pp. 1015-1027, Oct. 2004.
- [AS02] A. Arora, S. Subramaniam, „Wavelength Conversion Placement in WDM Mesh Optical Networks,” *Photonic Network Communication*, vol. 4, no. 2, pp. 167-177, 2002.
- [Aug02] S. Augsburger, *Using Dual-Supply, Dual-Threshold and Transistor Sizing to Reduce Power in Digital Integrated Circuits*, MSc Thesis, University of California, Berkley, 2002.
- [Bal09] A. Balandin, „Better Computing Through CPU Cooling,” *IEEE Spectrum*, Oct. 2009.
- [BG05] P. Becket, S. Goldstein, „Why area might reduce power in nanoscale CMOS,” *ISCAS*, pp. 2329-2332, May 2005.
- [BHM+02] R. Brodersen, M. Horowitz, D. Markovic et al., „Methods for true power minimization,” *ICCAD*, pp. 35-42, Nov. 2002.
- [BHM+02] R. Brodersen, M. Horowitz, D. Markovic, B. Nikolic et al., „Methods for True Power Minimization,” *ICCAD*, pp. 35-42, 2002.
- [Bis10] L. Bisdounis, „Short-circuit energy dissipation model for sub-100nm CMOS buffers,” *IEEE Int.Conf. on Electronics,Circuits and Systems*, pp. 615-618, Dec. 2010.
- [BMT+07] N. Brisebarre, J. Muller, A. Tisserand, S. Torres, „Hardware operators for function evaluation using sparse-coefficient polynomials,” *Electronics Letters*, vol. 42, no. 25, pp. 1441-1442, 2007.
- [BN02] D. Bitzaros, S. Nikolaidis, „Estimation of bit-level transition activity in data paths based on word-level statistics and conditional entropy,” *IEEE Proc. of Circuits DevicesSyst*, vol. 149, pp. 234-240, 2002.
- [BNW+04] B. Black, D. Nelson, C. Webb, N. Samra, „3D Processing Technology and Its Impact on iA32 Microprocessors,” *In Proc. of Int. Conf. on Computer Design*, pp. 316-318, 2004.
- [BNYT95] R. Burch, F. Najm, P. Yang, T. Trick, „A Monte Carlo approach for power estimation,” *IEEE Trans. on VLSI Systems*, vol. 1, No. 1, pp. 63-71, Mar 1995
- [BOI95] M. Borah, R. Owens, M. Irwin, „Transistor sizing for minimizing power consumption of CMOS circuits under delay constraint,” *ISLPED*, pp. 167-172, 1995.

- [Bor09] S. Borkar, „3D Integration for energy efficient system design,” *Sympos. on VLSI Technology*, pp. 58-59, June 2009.
- [BSK+01] K. Banerjee, S. Souri, P. Kapur, K. Saraswat, „3D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration,” *Proc. of IEEE*, vol. 89, no. 5, pp. 602-633, May 2001.
- [BTJ+99] E. Berglind, L. Thylen, B. Jaskorzynska, C. Svensson, „A comparison of dissipated power and signal-to-noise ratios in electrical and optical interconnects,” *Journ. of Lightwave Technol.*, vol. 17, no. 1, pp. 68-73, 1999.
- [CC05] B. Calhoun, A. Chandrakasan, „Ultra-Dynamic Voltage Scaling Using Sub-Threshold Operational and Local Voltage Dithering in 90nm CMOS,” *IEEE Journ. of Solid-State Circuits*, vol. 41, no. 1, pp. 238-245, Jan. 2005.
- [CCC+06] D. Chiou, S. Chen, C. Yeh, „Timing driven power gating,” *ACM Conf. on Design Automation*, pp. 121-214, 2006.
- [CCF03] D. Chen, J. Cong, Y. Fan, „Low Power High Level Synthesis for FPGA Architecture”, *Proc. of ISLPED*, pp. 134-139, Aug. 2003.
- [CEM99] A. Conn, M. Elfadel, W. Molzen, „Gradient based optimization of custom circuits using a static-timing formulation,” *In Proc. of DAC*, pp. 452-459, 1999.
- [CGC+06] J. Clarke, A. Gaffar, G. Constantinides, P. Cheung, „Fast word-level power models for synthesis of FPGA-based arithmetic,” *ISCAS*, pp. 1299-1302, Aug. 2006.
- [CGC05] J. Clarke, A. Gaffar, G. Constantinides, „Parametrized Logic Power Consumption Models for FPGA-based Arithmetic,” *Int. Workshop on Field-Programmable Logic and Applications*, pp. 626-629, Aug. 2005.
- [CHK+05] Y. Cheon, P. Ho, A. Kahng, S. Reda, Q. Wang, „Power-aware placement,” *ASP-DAC*, pp. 795-800, 2005.
- [Chr98] W. Chren, „One-hot residue coding for low-delay product CMOS design,” *Trans. on Circuits and Systems II*, vol. 45, no. 3, pp. 303-313, Mar. 1998.
- [Cir87] M. Cirit, „Estimating Dynamic Power Consumption of CMOS Circuits”, *ICCAD*, pp. 534-537, Nov. 1987.
- [CJL+98] Z. Chen, M. Johnson, L. Wei, W. Roy, „Estimation of Standby Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks,” *Int. Symp. on Low Power Electronics and Design*, pp. 239-244, Aug. 1998.
- [CJMP03] S. Choi, J. Jang, S. Mohanty, V. Prasanna, „Domain Specific Modeling for Rapid Energy Estimation of Reconfigurable Architectures”, *The Journ. of Supercomputing*, Vol. 26, No. 3, pp. 259-281, Nov. 2003.
- [CK88] H. Chen, S. Kang, „iCOACH: A Circuit Optimization Aid for CMOS High Performance Circuits,” *Proc. of ICCAD*, pp. 372-375, Nov. 1988.
- [CLL+00] K. Cao et al., „BSIM4 Gate Leakage Model Including Source-Drain Partition,” *Int. Electron Devices Meeting*, pp. 815-818, Dec. 2000.
- [CLT+10] S. Chen, R. Lin, H. Tung, K. Lin, „Power Gating Design for Standard-Cell-Like Structured ASICs,” *DATE*, pp. 514-519, Nov. 2010.
- [CN04] R. Conway, J. Nelson, „Improved RNS FIR filter architectures,” *IEEE Trans. on Circuits Systems II*, vol. 51, no. 1, pp. 26-28, Jan. 2004.
- [CRP] T. Chou, K. Roy, S. Prasad, „Estimation of Circuit Activity Considering Signal Correlations and Simultaneous Switching,” *In. Proc. ACM Int. Conf. on Computer-aided Design*, pp. 300-303, June 1994.
- [CRP94] T. Chou, K. Roy, S. Prasad, „Estimation of Circuit Activity Considering Signal Correlations and Simultaneous Switching”, *IEEE/ACM Conf. on CAD*, pp. 300-303, June 1994.
- [CWJ+98] Z. Chen, L. Wei, M. Johnson, K. Roy, „Estimation of standby leakage power in CMOS circuits considering accurate modeling of transistor stacks,” *IEEE Int. Conf. on Comput.-Aided Design*, pp. 10-12, Aug. 1998.
- [DCR03] S. Das, A. Chandrakasan, R. Reif, „Three-Dimensional Integrated Circuits: Performance, Design Methodology, and CAD Tools,” *ISVLSI*, pp. 13-18, 2003.
- [Den94] C. Deng, „Power analysis of CMOS/BiCMOS circuits”, *In Proc. of the Int. Workshop on Low Power Design*, pp. 3-8, Apr. 1994.
- [DMM+02] M. Donno, L. Macchiarulo, A. Macii, E. Macii, M. Poncino, „Enhanced Clustered Voltage Scaling for Low Power,” *Proc. of ACM Great Lakes Symposium on VLSI*, pp. 18-23, 2002.
- [DT05] V. Dagalahal, T. Tuan, „Methodology for High Level Estimation of FPGA Power Consumption,” *ASP-DAC*, pp. 657-660, Jan. 2005.
- [ED01] R.W. Erickson, D. Maksimovic, *Fundamentals of Power Electronics*, Springer 2001.

- [EJH+04] D. Elléouet, N. Julien, D. Houzet, J. Couzin, E. Martin, „Power Consumption Characterization and Modeling of Embedded Memories in Xilinx Virtex 400E FPGA,” *Euromicro Symposium on Digital System Design*, pp. 394-401, 2004.
- [EM09] M. Ercegovac, R. McIlhenny, „Design and FPGA implementation of radix-10 algorithm for square root with limited precision primitives,” *In Proc. on Asilom. Conf. on Sys., Sys., Comp.*, pp. 935-939, Nov. 2009.
- [ESJ06] D. Elléouet, Y. Savary, N. Joulieu, „An FPGA Power Aware Design Flow,” *PATMOS*, pp. 415-424, Sep. 2006.
- [FD85] J. Fishburn, A. Dunlop, „TILOS: A Posynomial Programming Approach to Transistor Sizing,” *Proc. of ICCAD*, pp. 326-328, Nov. 1985.
- [FLT+09] A. Florea, L. Li, M. Thornton, D. Matula, „A Discrete Logarithm Number System for Integer Arithmetic Modulo 2k: Algorithms and Lookup Structures,” *IEEE Trans. on Computers*, vol. 58, no. 2, Feb. 2009.
- [FP97] W. Freking, K. Parhi, „Low-power FIF digital filters using residue arithmetic,” *Conf. on Signals, Systems&Computers*, pp. 739-743, Nov. 1997.
- [FSB99] M. Fujita, A. Sakai, T. Baba, „Ultrasmall and ultralow threshold GaInAsP-InP microdisk injection lasers: design, fabrication, lasing characteristics, and spontaneous emission factor,” *IEEE Journ. of Selected Topics in Quantum Electronics*, vol. 5, no. 3, pp. 673-681, 1999.
- [GAF+07] S. Gupta, J. Anderson, L. Farragher, Q. Wang „CAD Techniques for Power Optimization in Virtex-5 FPGAs,” *CICC*, pp. 85-88, 2007.
- [GBC+01] S. Gunther, F. Binns, D. Carmen, J. Hall, „Managing the impact of Increasing Microprocessor Power Consumption,” *Intel Technology Journal Q1*, 2001.
- [GC08] K. Gbolagade, S. Cotofona, „An Efficient RNS to binary converter using the moduli set $\{2n+1, 2n, 2n-1\}$,” *Design of Circuits and Integ. Systems*, 2008.
- [GDK+07] A. Gupta, D. Dutt, F. Kurdahi, K. Khouri, M. Abadir, „LEAF: A System Level Leakage-Aware Floorplanner for SoCs,” *ASP-DAC*, pp. 274-279, Jan. 2007.
- [Geo94] B. George, „Power analysis of semi custom design”, *CICC*, pp. 249-252, 1994.
- [GG04] J. Gordy, R. Goubran, „A combined LPC based speech coder and filtered-XLMS algorithm for acoustic echo cancellation,” *IEEE Conf. on Acoustic, Speech and Signal Processing*, vol. 4, pp. 125-128, May 2004.
- [GSS08] L. O’Goroman, M. Sammon, M. Seul, *Practical algorithms for image analysis*, Cambridge University Press, New York, 2008.
- [GVC10] K. Gbolagade, G. Voicu, S. Cotofona, „An Efficient FPGA Design of Residue-to-Binary Converter for the Moduli Set $\{2n+1, 2n, 2n-1\}$,” *IEEE Trans. on VLSI Systems*, vol. 19, no. 8, pp. 1500-1503, June 2010.
- [HAI11] K. Haghdad, M. Anis, Y. Ismail, „Floorplanning for low power IC design considering temperature variations,” *Microelectronics Journal*, vol. 42, no. 1, pp. 89-95, 2011.
- [HCC+06] M. Haurylay, G. Chen, H. Chen et al., „On-Chip Optical Interconnect Roadmap: Challenges and Critical Directions,” *IEEE Journ. of Selected Topics in Quantum Electronics*, vol. 12, no. 6, pp. 1699-1705, 2006.
- [HDC11] H. Hussam, K. Dhamin, R. Come, „Static Power Estimation of CMOS Logic Blocks in a Library Free Design Environment”, *Int. Journ. of Design, Analysis and Tools for Circuits and Systems*, Vol. 1, No. 1, pp. 41-52, 2011.
- [HE+06] S. Henzler, G. Georgakos, M. Eireiner et al., „Dynamic state-retention flip-flop for fine-grained power gating with small design and power overhead,” *IEEE Journ. of Solid-State Circuits*, vol. 41, no. 7, pp. 1654-1661, July 2006.
- [HKQ+99] I. Hong, D. Kirovski, G. Qu et al., „Power Optimization of Variable-Voltage Core-Based Systems,” *IEEE Trans. on CAD of ICS*, vol. 18, no. 12, pp. 1702-1714, Dec. 1999.
- [HN97] J. Halter, F. Najm, „A gate-level leakage power reduction method for ultra-low-power CMOS circuits,” *CICC*, pp. 475-478, 1997.
- [HOK01] M. Hamada, Y. Ootaguro, T. Kuroda, „Utilizing surplus timing for power reduction,” *Proc. of the IEEE CICC*, pp. 89-92, May 2001.
- [HXV+04] W. Hung, Y. Xie, N. Vijaykrishnan, M. Kandemir, et al., „Total Power Optimization through Simultaneously Multiple-Vdd Multiple-Vth Assignment and Device Sizing with Stack Forcing,” *ISLPED*, pp. 144-149, 2004.
- [Ibm97] „Back to the Future: Copper Comes of Age,” *IBM Think Research*, no. 4, 1997.
- [IRTS] International Technology Roadmap for Semiconductors, <http://www.irts.net>
- [ISN04] F. Isihara, F. Sheikh, B. Nikolic, „Level conversion for dual-supply systems,” *IEEE Trans.*

- on *VLSI Systems*, vol. 12, no. 2, pp. 185-189, Feb. 2004.
- [JBK98] S. Jung, J. Baek, S. Kim, „Short-circuit power estimation of static CMOS circuits,” *Proc. of ASP-DAC*, pp. 545-549.
- [JC10] R. Jevtic, C. Carreras, „Power Estimation of Embedded Multipliers in FPGAs,” *IEEE Trans. on VLSI*, vol. 18, no. 5, pp. 835-839, May 2010.
- [JCC08] R. Jevtic, C. Carreras, G. Caffarena, „Fast and Accurate Power Estimation of FPGA DSP Components Based on High-level Switching Activity Models,” *Int. Journ. of Electronics*, vol. 95, no. 7, pp. 653-68, July 2008.
- [JD10] B. Jovanovic, M. Damjanovic, „Low power techniques for leakage power minimization,” *Proc. of 54th ETRAN*, pp. EL3.4-1-4, June 2010.
- [Jev09] R. Jevtic, *High-Level Power Estimation of DSP Circuits Implemented in FPGAs*, PhD thesis, Universidad Politécnica de Madrid, 2009.
- [JG04] R. Jejurikar, R. Gupta, „Dynamic Voltage Scaling for Systemwide Energy Minimization in Real-Time Embedded Systems,” *ISLPED*, pp. 78-81, 2004.
- [JJ10] B. Jovanovic, M. Jevtic, „FPGA Implementation of Throughput Increasing Techniques of the Binary Dividers,” *UNITECH*, pp. 397-401, Nov. 2010.
- [JC10] B. Jovanovic, R. Jevtic, C. Carreras, „Triple-bit method for power estimation of nonlinear digital circuits in FPGAs,” *Electronics Letters*, vol. 46, no. 13, pp. 903-905, June 2010.
- [JC11] R. Jevtic, B. Jovanovic, C. Carreras, „Power estimation of binary dividers implemented in FPGAs,” *Proc. of ACM Great Lakes Symposium on VLSI*, pp. 313-318, 2011.
- [JKS+99] G. Jochens, L. Kruse, E. Schmidt, W. Nebel, „A New Parameterizable Power Macro-Model for Datapath Components,” *DATE*, pp. 29-36, Jan. 1999.
- [JTB04] T. Jiang, X. Tang, P. Banerjee, „Macro-models for high level area and power estimation on FPGAs,” *Proc. of ACM Great Lakes Symposium on VLSI*, pp. 26-28, Apr. 2004.
- [KCH+01] J. Kolski, R. Chary, A. Henrold, B. Press, *Building Power Efficient PC*, Intel Press, 2001.
- [KOI92] S. Kim, R. Owens, M. Irwin, „Experiment with a performance driven layout generator,” *Proc. of DAC*, pp. 687-690, June 1992.
- [KR02] C. Kim, K. Roy, „Dynamic Vth scaling scheme for active leakage power reduction,” *DATE*, pp. 163-167, May 2002.
- [KR09] S. Katkoori, S. Roy, „A Framework for Power-Gating Functional Units in Embedded Microprocessor,” *IEEE Trans. on VLSI Systems*, vol. 17, no. 11, pp. 1640-1649, 2009.
- [Kri04] A. Krishnamoorthy, „Minimizing IC Power without Sacrificing Performance”, *EETimes*, May 2004.
- [KSM+98] T. Kuroda, K. Suzuki, S. Mita et al., „Variable Supply-Voltage Scheme for Low-Power High-Speed CMOS Digital Design,” *IEEE Journ. of Solid-State Circuits*, vol. 33, no. 3, pp. 454-462, Mar. 1998.
- [KTR07] I. Kuon, R. Tessier, J. Rose, „FPGA Architecture: Survey and Challenges,” *Foundation and Trends in Electronics Design Automation*, vol. 2, no. 2, pp. 135-253, 2007.
- [Kur04] V. Kursun, *Supply and Threshold Voltage Scaling Techniques in CMOS Circuits*, PhD thesis, University of Rochester, 2004.
- [KYT+02] T. Karnik, Y. Ye, J. Tschanz, L. Wei et al., „Total Power Optimization By Simultaneous Dual-Vt Allocation and Device Sizing in High Performance Microprocessors,” *Proc. of DAC*, pp. 486-491, 2002.
- [KZS+00] U. Keyser, H. Shumacher, U. Zeitler et al., „Fabrication of a single-electron transistor by current-controlled local oxidation of a two-dimensional electron system,” *Applied Physics Letters*, vol. 76, no. 4, Jan. 2000.
- [Lan94] P. Landman, *Low Power Architectural Design Methodologies*, PhD Thesis, University of California, Aug. 1994.
- [LCG+09] T. Lin, K. Chong, B. Gwee, J. Chang, „Fine-grained power gating for leakage and short-circuit power reduction by using asynchronous-logic,” *ISCAS*, pp. 3162-3165, May 2009.
- [LDB+04] D. Lee, H. Deogun, D. Blaauw, D. Sylvester, „Simultaneous State Vt and Tox assignments for total standby power minimization,” *DATE*, pp. 494-499, 2004.
- [LDB+04] D. Lee, H. Deogun, D. Blaauw, D. Sylvester, „Simultaneous State, Vt and Tox Assignment for Total Standby Power Minimization,” *DATE*, pp. 494-499, 2004.
- [Lew99] D. Lewis, „Complex Logarithmic Number System Arithmetic Using High-Radix Redundant CORDIC Algorithms,” *ARITH*, pp. 194-203, 1999.
- [LLC+05] H. Lee, K. Lee, Y. Choi, N. Chang, „Cycle-Accurate Energy Measurement and Characterization of FPGAs,” *Analog Integrated Circuits and Signal Processing*, vol. 42, no. 3, pp. 239-251, Mar. 2005.

- [LLW07] J. Lamoreux, G. Lemieux, S. Wilton, „GlitchLess: An Active Glitch Minimization Technique for FPGAs,” *In Proc. of Intern. Sympos on FPGAs*, pp. 156-165, Feb. 2007.
- [LLW08] J. Lamoreux, G. Lemieux, S. Wilton, „GlitchLess: Dynamic Power Minimization in FPGAs Through Edge Alignment and Glitching Filtering,” *IEEE Trans. on VLSI Systems*, vol. 16, no. 11, pp. 1521-1534, Nov. 2008.
- [LR95] P. Landman, J. Rabaey, „Architectural Power Analysis: The Dual Bit Type Method,” *IEEE Trans. on VLSI Systems*, vol. 3, no. 2, pp. 173-187, 1995.
- [LS00] S. Lee, T. Sakurai, „Run-time voltage hopping for low-power real-time systems,” *IEEE/ACM Proc. DAC*, pp. 806-809, June 2000.
- [MA11] N. Mehta, B. Amrutur, „Dynamic Supply and Threshold Voltage Scaling for CMOS Digital Circuits Using In-Situ Power Monitor,” *IEEE Trans. on VLSI*, vol. PP, no. 99, pp. 1-10, Apr. 2011.
- [Mac04] E. Macii, *Ultra Low-Power Electronics Design*, Kluwer Academic Publishers, Massachusetts, USA, 2004.
- [Mac08] F. Machado, *Switching Activity Analysis of Digital Electronic Circuits described at RTL using Probabilistic Techniques*, PhD Thesis, Technical University of Madrid, 2008.
- [Mak00] D. Maksimovic, *Logical Simulation – An Estimation of a limit properties of designed digital circuit*, PhD Thesis, Faculty of Electronic Engineering, University of Nis, 2000.
- [MDG93] J. Monteiro, S. Devadas, A. Ghosh, „Retiming sequential circuits for low power,” *ICCAD*, pp. 398-402, Nov. 1993.
- [MH01] C. Mulpuri, S. Hauck, „Runtime and Quality Tradeoffs in FPGA Placement and Routing,” *In Proc. of ACM Int. Symp. on FPGA*, pp. 29-36, 2001.
- [Mil96] D. Miller, „Optics for low-energy communications inside digital processors: quantum detectors, sources and modulators as efficient impedance converters,” *Optics Letters*, vol. 14, no. 2, pp. 146-148, 1996.
- [Min01] M. Minhas, *Iterative algorithms for Timing and Low-Power Driven VLSI Standard Cell Placement*, Master's thesis, University of Petroleum and Minerals, Saudi Arabia, 2001.
- [MNC+03] S. Mukhopadhyay, S. Neau, C. Cakici et al., „Gate leakage reduction for scaled devices using transistor stacking,” *IEEE Trans. on VLSI Systems*, vol. 11, no. 4, pp. 716-730, Sep. 2003.
- [MND+10] A. Molahosseini, K. Navi, C. Dadkhah, O. Kavehei, S. Timarchi, „Efficient reverse converter designs for the new 4-moduli sets $\{2n-1, 2n, 2n+1, 2(2n+1)-1\}$ and $\{2n-1, 2n+1, 2(2n), 2(2n)+1\}$ based on new CRTs,” *IEEE Trans on Circuits and Systems I*, vol. 57, no. 4, pp. 823-835, Apr. 2010.
- [MR04] H. Mahmoodi-Meimand, K. Roy, „Dual-Edge Triggered Level Converting Flip-Flops,” *IEEE ISCAS*, pp. 661-664, 2004.
- [MR04] H. Mahmoodi-Meimand, K. Roy, „Data-Retention Flip-Flops for Power-Down Applications,” *IEEE ISCAS*, pp. 677-680, 2004.
- [MRR05] S. Mukhopadyay, A. Raychowdhury, K. Roy, „Accurate Estimation of Total Leakage in Nanometer-Scale Bulk CMOS Circuits Based on Device Geometry and Doping Profile,” *IEEE Trans. on CAD of ICs and Systems*, vol. 24, no. 3, pp. 363-381, Mar. 2005.
- [MRT08] F. Machado, T. Riesgo, Y. Torroja, „Disjoint Region Partitioning for Probabilistic Switching Activity Estimation at RTL,” *PATMOS*, pp. 1145-1148, Sep. 2008.
- [MSN+04] D. Markovic, V. Stojanovic, B. Nikolic et al., „Methods for true energy-performance optimization,” *IEEE Journ. of Solid-State Circuits*, vol. 39, no. 8, pp. 1282-1293, Aug. 2004.
- [Moo65] G. Moore, „Cramming More Components onto integrated circuits,” *Electronics Magazine*, vol. 38, no. 8, pp. 114-117, 1965.
- [Nas11] S. Nassif, „Waiting for the Post-CMOS Godot,” *Keynote on Int. ACM Great Lakes Symposium on VLSI*, May 2011.
- [NBD+01] S. Nerendra, S. Borkar, V. De, D. Antoniadis, A. Chandrakasan, „Scaling of Stack Effect and its Application for Leakage Reduction,” *ISLPED*, pp. 195-200, 2001.
- [NDO+03] D. Nguyen, A. Davare, M. Orshansky, D. Chinnery et al., „Minimization of Dynamic and Static Power Through Joint Assignment of Threshold Voltages and Sizing Optimization,” *ISLPED*, pp. 158-163, Aug. 2003.
- [NHK+01] K. Nose, M. Hirabayashi, H. Kawaguchi, S. Lee, T. Sakurai, „Vth-hopping Scheme for 82% Power Savings in Low-voltage Processors,” *CICC*, pp. 93-96, 2001.
- [NM97] W. Nebel, J. Mermeet, *Low Power Design in Deep Submicron Electronics*, Kluwer Academic Publishers, Massachusetts, USA, 1997.
- [NS00] K. Nose, T. Sakurai, „Analysis and Future Trends on Short-circuit Power”, *IEEE Trans. on*

- CAD of IC and Systems*, Vol. 19, No. 9, pp. 1023-1030, 2000.
- [NS00a] K. Nose, T. Sakurai, „Analysis and Future Trends on Short-Circuit Power,” *IEEE Trans. on CAD of ICs and Systems*, vol. 19, no. 9, pp. 1023-1030, Sep. 2000.
- [NS00b] K. Nose, T. Sakurai, „Optimization of V_{dd} and V_{th} for Low-Power and High-Speed Applications,” *ASP-DAC*, pp. 469-474, Jun. 2000.
- [OIT02] Y. Ono, H. Inokawa, T. Takagashi, „Binary adder of multigate Single-Electron Transistors: specific design using Pass-Transistor Logic,” *IEEE Trans. on Nanotechnology*, vol. 1, no. 3, pp. 93-99, 2002.
- [Pap91] A. Papoulis, *Probability, random variables and stochastic processes*, McGraw Hill, New York, 1991.
- [Par99] B. Parhami, *Computer Arithmetic: Algorithms and Hardware Design*, Oxford University Press, Oxford, 1999.
- [PCK+08] H. Park, H. Che, W. Kim, Y. Kim, „High Performance Level-Converting Flip-Flop with a simple Pulse Generator and a Fast Latch,” *ITC-CSCC*, pp. 561-564, 2008.
- [PF09] V. Pavlidis, E. Friedman, *Three-dimensional integrated circuit design*, Morgan Kaufmann Publishers, 2009.
- [PGH+04] C. Piguet, J. Gautier, C. Heer, I. O'Connor, U. Schlichtmann, „Extremely Low-Power Logic,” *In Proc. of DATE*, pp. 656-661, 2004.
- [PHB06] K. Paulsson, M. Huebner, J. Becker, „On-line Optimization of FPGA Power Dissipation by Exploiting Run-time Adaption of Communication Primitives,” *Int. Symp. on Integrated Circuits and Systems Design*, pp. 283-288, Sep. 2006
- [Pig05] C. Piguet, *Low Power Electronics Design*, CRC Press, 2005.
- [PL11] S. Patil, D. Lilja, „A Programmable and Scalable Technique To Design Spintronic Logic Circuits Based On Magnetic Tunnel Junction,” *Int. ACM Great Lakes Symposium on VLSI*, pp. 7-12, May 2011.
- [PLS01] J. Pouwelse, K. Langendoen, H. Sips, „Dynamic Voltage Scaling on a Low-Power Microprocessor,” *MobiCom01*, pp. 251-259, 2001.
- [PSL06] C. Piguet, C. Schuster, J. Nagel, „Static and Dynamic Power Reduction by Architecture Selection,” *PATMOS*, pp. 659-668, July 2006.
- [PYW02] K. Poon, A. Yan, S. Wilton, „A flexible power model for FPGA,” *In Proc. FPL*, pp. 312-321, 2002.
- [Rab09] J. Rabaey, *Low Power Design Essentials*, Springer-Verlag, New York, 2009.
- [RCN03] J. Rabaey, A. Chandrakasan, B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, Prentice Hall, New Jersey, USA, 2003.
- [Riv10] F. Rivallon, „Reducing switching power with intelligent clock gating,” *EETimes*, June 2010.
- [RJD98] A. Raghunathan, N. Jha, S. Dey, *High-level Power Analysis and Optimization*, Kluwer Academic Publishers, Massachusetts, USA, 1998.
- [RK97] R. Rogenmoser, H. Kaeslin, „The impact of transistor sizing on power efficiency in submicron CMOS circuits,” *IEEE Journ. of Solid-State Circuits*, vol. 32, no. 7, pp. 1142-1145, July 1997.
- [RMM03] K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, „Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-submicrometer CMOS circuits,” *Proc. of the IEEE*, vol. 91, no. 2, pp. 305-327, Feb. 2003.
- [RSH97] S. Ramprasad, N. Shanbhag, I. Hajj, „Analytical Estimation of Signal Transition Activity from Word-Level Statistics,” *IEEE Trans. on CAD*, vol. 16, no. 7, pp. 718-733, Mar. 1997.
- [RWM85] D. Radhakrishnan, S. Whittaker, G. Maki, „Formal Design Procedures for Pass-Transistor Switching Circuits,” *IEEE Journ. of Solid-State Circuits*, vol. SC-20, no. 2, pp. 531-536, Apr. 1985.
- [Sal10] B. Salami, „Chip-dependent leakage Power-Aware Placement algorithm for FPGAs,” *ELECO*, pp. 350-360, 2010.
- [Sch95] P. Schneider, „PAPSAS: A fast switching activity simulator”, *PATMOS'95*, pp. 351-360, 1995.
- [SD09] G. Sutter, J. Deschamps, „High-speed fixed point dividers for FPGAs,” *In Proc. FPL*, pp. 448-452, Aug. 2009.
- [SDB+04] G. Sutter, J. Deschamps, G. Bioul, E. Boemo, „Power aware dividers in FPGAs,” *PATMOS*, pp. 574-584, Aug. 2004.
- [SJ11] N. Stamenkovic, B. Jovanovic, „Reverse Converter Design for the 4-Moduli Set $\{2n-1, 2n, 2n+1, 2(2n+1)-1\}$ Based on the Mixed-Radix Conversion,” *Facta Universitatis ser.*

- Electronics and Energetics*, vol. 24, no. 1, pp. 91-105, Apr. 2011.
- [SKB02] L. Shang, A. Kaviani, K. Bathala, „Dynamic Power Consumption in Virtex-II FPGA Family,” *ACM/SIGDA Int. Symp. on FPGAs*, pp. 157-164, Feb. 2002.
- [SN90] T. Sakurai, R. Newton, „Alpha-power law MOSFET model and applications to CMOS inverter delay and other formulas,” *IEEE Journal on Solid-State Circuits*, vol. 25, no. 2, pp. 584-594, 1990.
- [SNP+06] C. Shuster, J. Nagel, C. Piguet, P. Farine, „Architectural and Technology influence on the Optimal Total Power Consumption,” *DATE*, pp. 13-19, 2006.
- [Sor06] N. Sorokin, „Implementation of high-speed fixed-point dividers on FPGA,” *Journal of CS&T*, vol. 6, no.1, pp. 8-11, Apr. 2006.
- [SP00] J. Sataynarayana, K. Parhi, „Theoretical Analysis of Word-Level Switching Activity in the Presence of Glitching and Correlation,” *IEEE Trans. on VLSI Systems*, vol. 8, no. 2, pp. 148-159, Apr. 2000.
- [SPN+05] C. Shuster, C. Piguet, J. Nagel, P. Farine, „An Architecture Design Methodology for Minimal Total Power Consumption at Fixed Vdd and Vth,” *Journ. of Low Power Electronics*, vol. 1, no. 1, pp. 3-10, Apr. 2005.
- [SPV+11] S. Srinivasan, B. Phanibhapphana, A. Vijayakumar, S. Kundu, „Stress Aware Switching Activity Driven Low Power Design of Critical Paths in Nanoscale CMOS circuits,” *Int. Great Lakes Symposium on VLSI*, pp. 265-269, May 2011.
- [Sri00] A. Srivasta, „Back gate bias method of threshold voltage control for the design of low voltage CMOS ternary logic,” *Microelectronics Reliability*, vol. 40, no. 3, pp. 2107-2110, 2000.
- [SRV+93.] S. Sapatnekar, V. Rao, P. Vaidya, S. Kang, „An Exact Solution to the Transistor Sizing Problem for CMOS circuits using Convex Optimization,” *IEEE Trans. on CAD*, vol. 12, no. 11, pp. 1621-1634, Nov. 1993.
- [SS02] R. Shelar, S. Sapatnekar, „An Efficient Algorithm for Low Power Pass Transistor Logic Synthesis,” *ASP-DAC*, pp. 87-92, Jan. 2002.
- [SS03] A. Srivastava, S. Sylvester, „Minimizing Total Power by Simultaneous Vdd/Vth Assignment,” *ASP-DAC*, pp. 400-403, Jan. 2003.
- [SS08] K. Siozios, D. Soudris, „A Power-Aware Placement and Routing Algorithm Targeting 3D FPGAs,” *Journ. of Low-Power Electronics*, vol. 4, no. 3, pp. 275-289, Dec. 2008.
- [SSB04] A. Srivastava, D. Sylvester, D. Blaauw, „Concurrent Sizing, Vdd and Vth Assignment for Low-Power Design,” *DATE*, pp. 718-719, Feb. 2004.
- [SSK+87] B. Sheu, D. Schafretter, P. Ko, M. Jeng, „BSIM: Berkley shorg-channel IGFET model for MOS transistors,” *IEEE Journ. of Solid-State Circuits*, vol. 22, no. 4, pp. 558-566, Aug. 1987.
- [TB05] E. Todorovich, E. Boemo, „Statistical Power Estimation for FPGAs,” *LNCS*, pp. 340-349, Jan. 2005.
- [TB05] E. Todorovich, E. Boemo, „Statistical Power estimation for FPGAs”, *LNCS*, pp. 515-518, June 2005.
- [TGS+02] E. Todorovich, M. Gilabert, G. Sutter et al., „A Tool for Activity Estimation in FPGAs,” *LNCS*, pp. 340-349, June 2002.
- [TKT05] N. Takagi, S. Kadowaki, K. Takagi, „A Hardware Algorithm for Integer division,” *IEEE Symp. on Computer Arithmetic*, pp. 140-146, 2005.
- [Tod02] E. Todorovich, „A tool for activity estimation in FPGAs”, *LNCS*, pp. 340-349, June 2002.
- [Tod06] E. Todorovich, *Statistical Power Estimation in FPGAs*, PhD thesis, Universidad Autónoma de Madrid, 2006.
- [UI00] K. Usami, M. Igarashi, „ Low-Power Design Methodology and Applications utilizing Dual Supply Voltages,” *ASP-DAC*, pp. 123-128, 2000.
- [Vee84] H. Veendrick, „Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits,” *IEEE J.Solid-State Circuits*, vol. sc-19, no. 4, pp. 468-473, Aug. 1984.
- [VKP+10] K. Vorwerk, A. Kennings, V. Pevzner, A. Kundu et al., „Power minimization during field programmable gate array placement,” *IET Computer&Digital Techniques*, vol. 4, no. 3, pp. 170-183, May 2010.
- [VS94] S. Vemuru, N. Scheinberg, „Short-circuit power dissipation estimation for CMOS logic gates,” *IEEE Trans. on Circuits Systems*, vol. 41, no. 3, pp. 762-765
- [WBG+06] R. Wain, I. Bush, M. Guest, M. Deegan et al., „An Overview of FPGAs ans FPGA programming: Initial experiences at Daresbury, CCLRC Daresbury Laboratory,

- Warrington, Cheshire, Nov. 2006.
- [WV98] Q. Wang, S. Vrudhula, „On short-circuit power estimation of CMOS inverters,” *Proc. of ICCD*, pp. 70-75, Oct. 1998.
- [Xil05] Xilinx. Xilinx University Program Virtex-II Pro Development System. Hardware Reference Manual, v1.0, Mar. 2005.
- [Xil06] Xilinx IP Core Divider Datasheet, http://www.xilinx.com/support/documentation/ipmath_dividereciprocalfunction_divider.htm, 2006.
- [Xil11] Xilinx Xcell Journal, Third Quarter 2011, Issue 76, pp. 8-15, 2011.
- [YBD98] Y. Ye, S. Borkar, V. De, „A Technique for Standby Leakage Reduction in High-Performance Circuits,” *Symp. of VLSI circuits*, pp. 40-41, 1998.
- [YVC+97] I. Yang, C. Vieri, A. Chandrakasan, D. Antoniadis, „Back-Gated CMOS on SOIAS For Dynamic Threshold Voltage Control,” *IEEE Trans. on Electron Devices*, vol. 44, no. 5, pp. 822-831, May 1997.

Апстракт

Током протеклих деценија усавршавање технолошког процеса производње CMOS интегрисаних кола незадрживо се одвијало. Према чувеном Муровом закону, темпо овог развоја био је невероватан: на сваке две године број транзистора које је било могуће имплементирати у једно интегрисано коло отприлике се удвостручио. Ово континуирано скалирање CMOS технолошког процеса одвело нас је дубоко у под-микронске режиме у којима ограничавајући фактор није више функционалност на чипу већ потрошња електричне енергије. Бројни портабилни батеријски напајани уређаји са невероватним перформансама доказ су ове тврдње. Почев од 180nm технолошких процеса па надаље статичка потрошња електричне енергије, последица неактивних „цурећих“ транзистора, постаје незанемарљив извор дисипације електричне енергије. Стога приликом оптимизација потрошње дизајна постаје неопходно третирати је равноправно са динамичком потрошњом.

Први део ове дисертације бави се проблемима оптимизације тј. минимизације потрошње (како статичке тако и динамичке) електричне енергије. Највеће уштеде у потрошњи електричне енергије дизајна могуће је остварити на системском нивоу тока његовог пројектовања на коме се дефинишу архитектура и алгоритми дизајна. Утицати на потрошњу а при том не жртвовати перформансе система постаје све теже и теже на нижим нивоима пројектовања система. Селектовање најефикаснијег (у погледу потрошње електричне енергије) алгоритма из групе расположивих захтева могућност брзе и тачне процене потрошње сваке имплементације. Овакве процене омогућиле би да се избегну временски захтевне физичке имплементације сваке могуће архитектуре дизајна.

Стога је други део дисертације оријентисан ка моделима високог нивоа за процену динамичке потрошње логике DSP (Digital Signal Processing) дизајна. Овај део дисертације представља наставак рада на моделима за процену потрошње описаног у [Jev09] и описује методологију за процену динамичке потрошње логике високог нивоа, базирану на логичкој структури компоненте и аналитичком прорачуну транзиционе активности која се јавља унутар компоненте. Аналитички модели за процену транзиционе активности (динамичке потрошње) унутар логичке структуре дизајна су параметризовани а параметри на основу којих су у стању да процене вредност динамичке потрошње су величина улазних сигнала операнада (број битова) као и њихова статистика (средња вредност, варијанса, и аутокорељација).

Најпре је представљено побољшање већ постојећег модела за процену транзиционе активности (динамичке потрошње) логичке структуре бинарног множача – компоненте која на својим излазима испољава одређену врсту

нелинеарности која у претходним моделима за процену динамичке потрошње није узимана у обзир. Експерименталним путем утврђено је да модел предложен у дисертацији даје значајно тачније (у неким случајевима 4-5 пута) процене динамичке потрошње.

Након тога предложен је аналитички модел за процену динамичке потрошње логичке структуре бинарног делитеља. У литератури која је била на располагању, аутор дисертације није наишао на сличне покушаје процене динамичке потрошње.

Ради евалуације предложених модела DSP дизајни имплементирани су на Virtex II Pro FPGA компоненту XUP (Xilinx University Program) развојне плоче. Поред измерених вредности динамичке потрошње, као референца за евалуацију модела коришћен је и XPower Analyzer алат у оквиру Xilinx ISE програмског пакета. Средње релативне грешке модела за процену динамичке потрошње рачунане у односу на измерене и XPower вредности динамичке потрошње не прелазе 10%, што је јако добар резултат имајући у виду чињеницу да су модели на високом нивоу апстракције система. Поред тога што су задовољавајуће тачности, модели за процену потрошње су изузетно брзи. Процене динамичке потрошње добијају се за време реда десетине секунде.

Abstract

During the past years the progress of silicon process technology marched on relentlessly. According to Moore's law, silicon process technology continues improvement at an astonishing pace. Every 2 years the number of transistors that can be integrated on a single IC approximately doubles. The continued scaling of the CMOS technology has led us into the deep submicron regimes where design is not limited by the functionality on a chip but is constrained with its power consumption. Numerous portable battery-powered electronic devices with outstanding performances are the prove of previous statement. Starting from 180nm technologies, static power consumption due to leaky "off" transistors is now a non negligible source of power dissipation, even in running mode. Thus, the total power consumption (i.e. dynamic plus static power) has to be optimized instead of simply reducing dynamic power.

The first part of this work is oriented toward the problems of power consumption (both static and dynamic) minimization techniques. The largest impact on power reduction can be achieved at the system level where architecture and algorithms are to be defined. Influencing power consumption without sacrificing system performances is becoming harder and harder at the lower levels of system design flow. Selecting the most power efficient algorithm out of many available requires a fast and accurate way to estimate the power consumption of any implementation. In this way, time-consuming low-level implementations of each possible design architecture will be avoided.

Therefore, the second part of this thesis deals with high-level dynamic power estimation models of DSP-oriented (Digital Signal Processing) designs. This dissertation is the continuation of the work on power estimation models described in [Jev09] and presents the methodology for high-level logic power estimation which is based on the component's structure and the analytical computation of the switching activity produced inside the component. Power estimation models are parametrized in terms of input signal statistics (mean, variance, autocorrelation) and operands' word-lengths.

The improvement of the existing model for *ripple carry array* multiplier switching activity (dynamic power) estimation is firstly presented. Unlike previous DBT (Dual-Bit Type) model, the TBT (Triple-Bit Type) model presented here takes into account non-linearities in bit-level switching activities occurring at the multiplier output. It is experimentally confirmed that the model proposed in this thesis gives far better power estimations (four to five times) than the previous one.

After that, the methodology for high-level logic power estimation of binary divider is presented. To the best of author's knowledge no previous work focuses on high-level divider power estimation.

In order to evaluate proposed models, DSP designs are implemented on Virtex II Pro FPGA device on XUP (Xilinx University Program) board and dynamic on-board power

consumption is measured. Apart on-board dynamic power, as a reference for our models evaluation, XPower Analyzer tool within the Xilinx ISE software package is used. The mean relative errors of our high-level power estimation models are less than 10% which is very encouraging result, given a fact that the models have high-level nature. In addition to high accuracy, proposed methodology for dynamic power estimation is very fast. The power estimates are obtained in order of milliseconds.

Биографија

Бојан Јовановић рођен је 01.01.1981. године у Лесковцу. Основну школу и природно-математички смер Гимназије „Стојан Љубић“ завршио је у Лебану са просечним оценама 5.00, као носилац Вукових диплома.

Школске 2000/2001. године почиње студије на Електронском факултету Универзитета у Нишу, смер Електроника, и завршава их септембра 2006. године, са просечном оценом 9.30 и оценом 10 на дипломском испиту. За време студија два пута је, од стране факултета, награђиван за постигнуте резултате током студирања. Школске 2006/2007. године уписује докторске студије на Електронском факултету у Нишу, студијски програм Електроника. Све испите предвиђене планом и програмом положио је са просечном оценом 10.

Професионално искуство започиње 2007. године као инжењер конструктор електронских уређаја у фирми „Сентронис“ А.Д. са седиштем у Нишу. У фирми је радио на пословима производње и развоја сензора за мерење струје и магнетног поља базираних на Hall-овом ефекту, пословима аутоматизације процеса калибрације магнетних сензора уз помоћ LabVIEW програмског пакета, развоја ручног Тесламетра ниске цене, као и на пословима развоја дигиталног дела Hall-овог Тесламетра који мери магнетно поље у простору (x, y, z компоненте поља).

Од априла 2008. године запошљен је на Електронском факултету у Нишу, најпре као сарадник у настави а након тога, од марта 2011. године, и као асистент на катедри за Електронику. Ангажован је на пословима извођења аудитивних и лабораторијских вежби из предмета Дигитална електроника. Реализовао је лабораторијске вежбе базирание на развојној плочи Altera DE1 за предмет Дигитална алектроника, а коаутор је и Практикума за лабораторијске вежбе из истог предмета. У периоду од јануара до краја марта 2010. године боравио је на Политехничком Универзитету у Мадриду у оквиру међународне размене истраживача међународног Tempus пројекта *System-on-chip design*. Као аутор или коаутор објавио је 22 научна рада. Радови су објављени у међународним или домаћим научним часописима и на међународним или домаћим научним конференцијама и симпозијумима.

Члан је ACM и IEEE друштва. Активно се служи енглеским, а пасивно француским језиком.