

УДК 681.3

Б.Р. Дунець, А.О. Мельник

Національний університет "Львівська політехніка",
кафедра "Електронні обчислювальні машини"**ДОСЛІДЖЕННЯ ХАРАКТЕРИСТИК КОНТРОЛЕРА ДОСТУПУ
ДО БАГАТОБЛОКОВОЇ ПАМ'ЯТІ**

© Дунець Б.Р., Мельник А.О., 2002

Запропоновано нарощувану архітектуру контролера доступу до багатоблокової пам'яті для безконфліктного доступу процесорів багатопроцесорної системи до блоків спільної пам'яті. Визначено основні характеристики контролера та їх залежність від вхідних значень параметризації.

The scalable architecture of Multiblock RAM Access Controller is proposed. Controller is intended for performing conflict-free access of microprocessors to blocks of shared memory. Main controller characteristics and its relation with parameters of scaling are determined.

Вступ

Одним із компонентів спеціалізованої багатопроцесорної системи (СБС) [1] є контролер доступу (КД) до спільної багатоблокової пам'яті. Параметризована модель СБС [2] призначена для використання в засобах високорівневого проектування [3] спеціалізованих надвеликих інтегральних схем. Завдяки застосуванню спільної багатоблокової пам'яті досягається максимальна продуктивність СБС, оскільки є можливим одночасний доступ декількох процесорних елементів до блоків спільної пам'яті.

Автоматичний синтез КД вимагає наявності базової параметризованої моделі та спеціалізованого генератора. Досягнення максимальної продуктивності СБС можливе при оптимальному виборі параметрів синтезу КД, при яких досягаються необхідні характеристики його роботи. Оптимальний вибір таких параметрів наштовхується на ряд проблем, оскільки характеристики контролера залежать не тільки від структурних параметрів КД, але і від характеру вхідного потоку даних. У статті досліджено статичні та динамічні складові характеристик КД, їх залежність від структурних параметрів та характеру вхідного потоку даних, наведено методику їх обчислення.

1. Базова архітектура контролера

Контролер доступу до багатоблокової пам'яті забезпечує обслуговування N вхідних запитів A_i ($i=1,2,\dots,N$) до M блоків R_j ($j=1,2,\dots,M$) пам'яті (рис.1). Кожен із вхідних запитів A_i містить інформацію про номер блоку пам'яті, адресу комірки в даному блоці, тип запиту (читання чи запис) та вхідні дані (якщо тип запиту – запис). КД за час T обслуговує запити, тобто записує та читає інформацію в/з блоків пам'яті та видає прочитані дані (для запитів типу "читання") на відповідних вихідних лініях даних D_i ($i=1,2,\dots,N$). У випадку перевищення кількості вхідних запитів допустимої величини виникає переповнення КД та встановлюється вихідний інформаційний сигнал *Overflow* в '1'.

Необхідно зазначити, що КД є повністю синхронним пристроєм, тому вхідні запити, що надійшли в один момент часу, обслуговуються одночасно. Особливістю роботи КД є

наявність фіксованого часу обслуговування T – проміжку часу між моментом надходження запиту на читання та моментом видачі прочитаних даних. Так як КД є повністю синхронним пристроєм, то будемо розглядати його функціонування в дискретному часі $t=0,1,\dots$, де один проміжок часу дорівнює періоду тактового імпульсу.

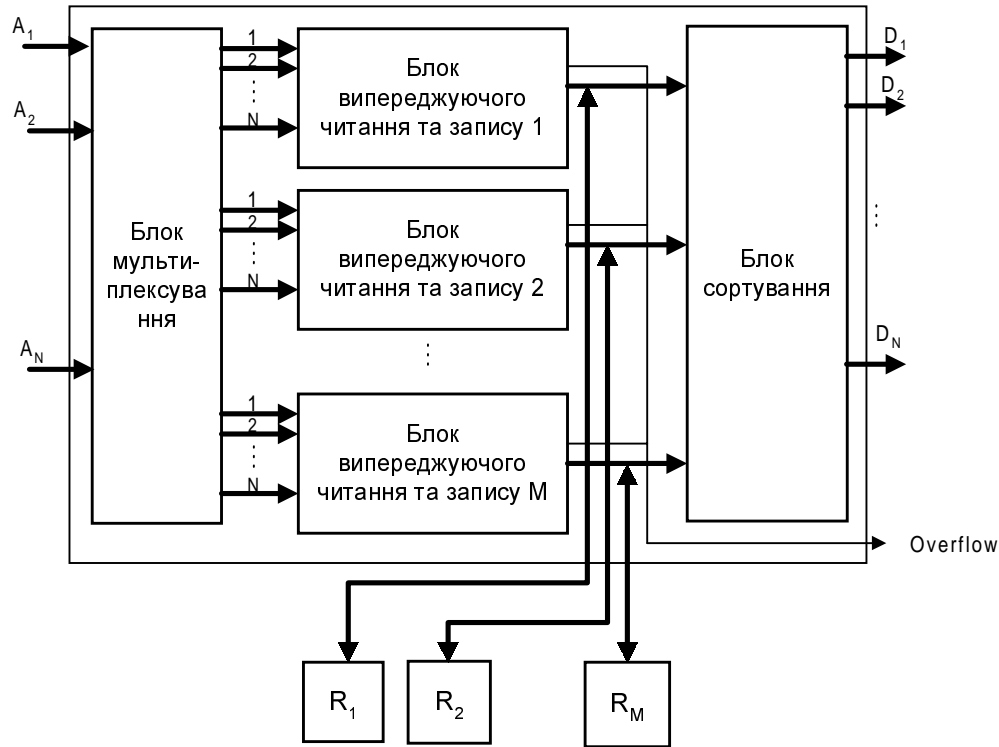


Рис. 1. Архітектура контролера

Контролер доступу до багатоблокової пам'яті складається з блоку мультиплексування, блоку сортування та M ідентичних блоків випереджувального читання та запису. Блок мультиплексування виконує поділ вхідної послідовності запитів на M підмножин запитів, кожна з яких містить тільки запити, що належать до одного блоку пам'яті. Кожен з блоків випереджувального читання та запису обслуговує запити, що стосуються тільки одного блоку пам'яті. Блоки випереджувального читання та запису обслуговують запити, тобто читають інформацію з відповідної комірки пам'яті чи записують дані в задану комірку пам'яті. Прочитані дані разом зі службовою інформацією надходять у блок сортування, який впорядковує прочитані дані з метою видачі даних, що відповідають вхідним запитам в точно визначений момент часу на відповідному виході.

Запропоновану архітектуру КД вирізняє ряд суттєвих властивостей:

1. Запити, що належать до різних блоків пам'яті, є незалежними. Кількість запитів $W(t)$ до КД у момент часу t можна визначити як суму кількості запитів до різних блоків пам'яті:

$$W(t) = \sum_{j=1}^M V_j(t), \quad (1)$$

де $V_j(t)$ – кількість запитів до j -го блоку пам'яті у момент часу t .

2. Блок випереджувального читання та запису дозволяє приймати, накопичувати та обслуговувати запити, що надійшли в КД не тільки в даний момент часу, а і протягом деякого заданого фіксованого періоду часу F , вікна обслуговування.

3. Оскільки час обслуговування запитів $T(t)$, що надійшли в КД у момент часу t , залежить не тільки від структури, а і від характеру вхідного потоку запитів, то можна розділити динамічні та статичні характеристики КД.

4. Основними параметрами, які впливають на статичні та динамічні параметри КД, є: N – кількість вхідних запитів, M – кількість блоків пам'яті, F – значення вікна обслуговування, а також розрядність шини даних та адреси.

2. Статичні характеристики КД

Статичні характеристики КД не залежать від характеру вхідних даних, а визначаються структурою пристрою. До таких характеристик КД можна віднести: мінімальний час обслуговування T_{\min} та апаратну складність пристрою.

Важливим параметром, від якого залежить час обслуговування запитів $T(t)$, є *мінімальний час* T_{\min} обслуговування вхідних запитів. Мінімальний час обслуговування (кількість тактів) визначається затримкою конвеєра (кількістю конвеєрних реєстрів) тракту обслуговування КД:

$$T_{\min} = N_{\text{MUX}} + N_{\text{SRW}} + N_{\text{SORT}}, \quad (2)$$

де N_{MUX} – кількість конвеєрних реєстрів блоку мультиплексування, $N_{\text{SRW}} = F + 1$ – кількість конвеєрних реєстрів блоку випереджувального читання та запису, N_{SORT} – кількість конвеєрних реєстрів блоку сортування.

Наприклад, при $F=4$, $N_{\text{MUX}} = 0$ (комбінаційна схема) та $N_{\text{SORT}} = 1$, мінімальний час обслуговування дорівнює $T_{\min} = 0 + (4 + 1) + 1 = 6$ тактів.

Якщо кількість вхідних запитів не перевищує допустиму межу, то запити, що надійшли в КД в один момент часу, обслуговуються КД за чітко визначений час, що дорівнює T_{\min} .

3. Динамічні характеристики КД

Розгляд динамічних параметрів обумовлено виникненням переповнення КД ($\text{Overflow} = '1'$). Таке переповнення виникає при перевищенні вхідних запитів допустимої кількісної межі, що веде до припинення прийому нових запитів, обслуговування вже накопичених та до збільшення часу обслуговування. Динамічні характеристики КД залежать від характеру вхідних даних. До таких характеристик КД можна віднести: час обслуговування запитів, що надійшли в момент часу t до КД $T(t)$, кількість додаткових тактів на обслуговування запитів, що надійшли в момент часу t до КД $TA(t)$, загальна кількість тактів TA очікування.

Оскільки запити, що належать до різних блоків пам'яті, є незалежними, то проведемо аналіз впливу параметрів на динамічні характеристики КД для одного блоку пам'яті з подальшим узагальненням для КД загалом.

Оскільки в КД використано випереджувальне обслуговування запитів, то час обслуговування запитів залежить не тільки від кількості запитів $V_j(t)$, що надійшли до j -го блоку пам'яті у даний момент часу, але й від кількості запитів, які надійшли в пристій та знаходяться в обслуговуванні.

Коефіцієнт заповнення $k_j(t)$ є величиною, яка вказує на кількість накопичених запитів до j -го блоку пам'яті в момент часу t .

Теорема. При $k_j(t) \leq F$, час обслуговування запитів до j -го блоку пам'яті, що надійшли в момент часу t , дорівнює $T_j(t) = T_{\min}$.

Доведення. Блок випереджувального читання та запису в один момент часу може звернутись тільки до однієї комірки пам'яті, тому загальна продуктивність КД обмежується пропускнуою здатністю пам'яті. Відповідно, за період часу F контролер може обробити F запитів, тоді при кількості запитів $k_j(t) \leq F$ переповнення не виникає.

Наслідок. Якщо $k_j(t) > F$, то кількість запитів перевищує критичну межу і КД необхідно один такт часу на обслуговування кожного додаткового запиту. Оскільки прийнята кількість запитів в КД не дозволяє приймати нові, то КД інформує зовнішні пристрої про переповнення сигналом $Overflow='1'$ та виконує обслуговування існуючих запитів до моменту зникнення переповнення.

Кількість додаткових тактів на обслуговування запитів $TA_j(t)$, що надійшли в момент часу t до j -го блоку пам'яті:

$$\text{Якщо } k_j(t) > F, \text{ то } TA_j(t) = k_j(t) - F, \quad (3)$$

тоді загальний час обслуговування запитів $T_j(t)$ у момент часу t до j -го блоку пам'яті збільшується на величину кількості додаткових тактів за період F :

$$T_j(t) = T_{\min} + \sum_{i=0}^{F-1} TA_j(t-i) \quad (4)$$

Коефіцієнт заповнення $k_j(t)$ та кількість додаткових тактів на обслуговування запитів $TA_j(t)$ визначаються в такій послідовності:

1. $k_j(t < 0) = 0$, $TA_j(t < 0) = 0$;
2. Якщо $k_j(t-1) = 0$, то $k_j(t) = k_j(t-1) + V_j(t) - TA_j(t-1)$ інакше $k_j(t) = k_j(t-1) + V_j(t) - TA_j(t-1) - 1$;
3. Якщо $k_j(t) > F$, то $TA_j(t) = k_j(t) - F$, інакше $TA_j(t) = 0$.

Приклад для $F=3$, $T_{\min}=5$

T	10	9	8	7	6	5	4	3	2	1	0
$V_j(t)$	0	0	1	4	0	0	1	4	3	0	1
$k_j(t)$	1	2	3	4	1	2	3	6	3	0	1
$TA_j(t)$	0	0	0	1	0	0	0	3	0	0	0
$T_j(t)$	5	6	6	6	5	8	8	8	5	5	5

Мінімальний час обслуговування КД загалом не залежить від кількості блоків пам'яті M , оскільки структура блоків випереджувального читання та запису ідентична.

Кількість додаткових тактів на обслуговування запитів $TA(t)$ визначається як максимальне значення серед кількості додаткових тактів обслуговування кожного з блоків, оскільки при переповненні одного із блоків випереджувального читання та запису інші продовжують обслуговувати накопичені запити:

$$TA(t) = \max_{j=1 \dots M} TA_j(t). \quad (5)$$

Відповідно,

$$T(t) = T_{\min} + \sum_{i=0}^{F-1} TA(t-i) \quad (6)$$

Загальна кількість тактів ТА очікування при обслуговуванні КД послідовність запитів за деякий період часу надходження вхідних запитів T' дорівнює

$$TA = \sum_{t=0}^{T'-1} TA(t) \quad (7)$$

КД до багатоблокової пам'яті, в якому не використовується алгоритм випереджувального запису та читання, обслуговує тільки поточні запити, не використовуючи їх накопичення чи перестановки. Відповідно, час обслуговування запиту такого контролера залежить від кількості запитів до одного блоку пам'яті в поточний момент часу i не залежить від попередніх запитів ($F=1$, $T_{\min}=1$), тобто $TA_j(t) = V_j(t) - 1$; $T_j(t) = T_{\min} + TA_j(t) = 1 + TA_j(t)$

4. Визначення числових значень характеристик контролера

Оскільки динамічні характеристики КД залежать від вхідного потоку даних, то на основі імовірнісних характеристик вхідної послідовності запитів моделюванням можна визначити конкретні числові значення характеристик КД визначеної структури. Завдяки запропонованій методиці можна оцінити характеристики КД без синтезу його структури, проводячи тільки функціональну симуляцію. У протилежному випадку необхідно було б синтезувати КД та визначати його характеристики на основі симуляції роботи КД на структурно-функціональному рівні (напр. VHDL), що є тривалим процесом.

При порівнянні контролерів із різними структурами основним критерієм, який характеризує продуктивність, є коефіцієнт простою:

$$E = \frac{TA}{T' + TA}, \quad (8)$$

який визначається як співвідношення загальної кількості тактів очікування за час роботи пристрою до загального часу обслуговування.

Оскільки характеристики КД залежать від характеру вхідного потоку, то основною величиною, яка характеризує вхідний потік запитів, є коефіцієнт інтенсивності надходження запитів:

$$I = \frac{W}{N \cdot T'}, \quad (9)$$

який визначається як співвідношення кількості запитів, що надійшли в КД за час роботи пристрою до максимально можливої кількості запитів за цей же період.

З використанням запропонованої методики було проведено симуляцію та отримано графіки залежності (рис. 2) простою E від інтенсивності надходження запитів I для різних типів контролерів для фіксованих значень $N=10$, $M=5$ та рівноімовірного звертання до блоків пам'яті.

Як бачимо з рис. 2, коефіцієнт простою КД, в якому не використовується алгоритм випереджувального читання та запису (крива Simple Controller), є завжди більший, ніж в контролерів доступу до багатоблокової пам'яті, в яких використовується алгоритм випереджувального читання та запису.

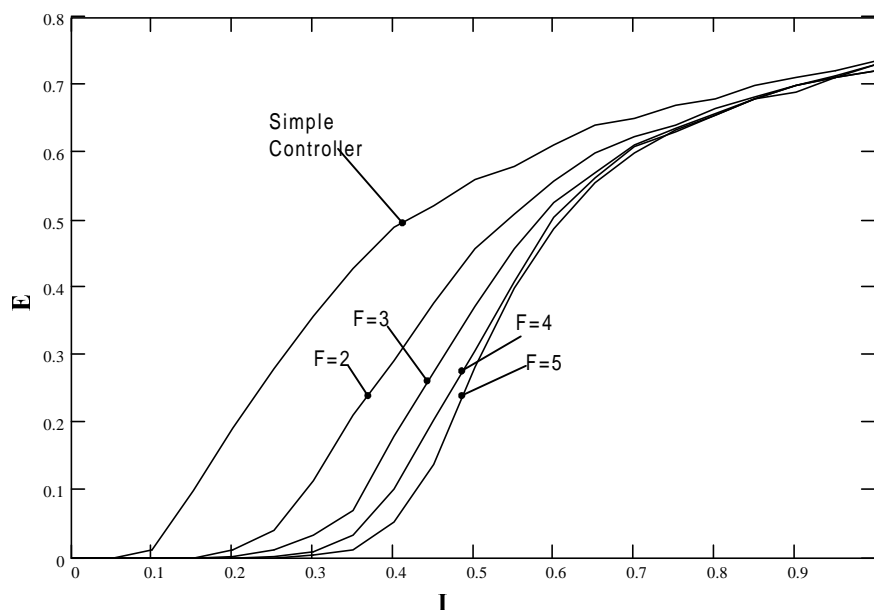


Рис. 2. Ефективність роботи КД

Висновки

Розглянута архітектура контролера доступу до багатоблокової пам'яті дозволяє побудувати спеціалізований КД з довільною кількістю вхідних портів, блоків пам'яті та тактів очікування. Запропонована методика дозволяє визначити характеристики КД на основі вхідних параметрів та характеристики вхідної послідовності. Показано, що запропонований КД, в якому використовується алгоритм випереджувального читання та запису, є більш ефективний порівняно з аналогічними КД.

1. *High-level synthesis of nonprogrammable hardware accelerators.* R.Schreiber, S. Aditya, B.R. Rau, V. Kathail, S. Mahlke, S. Abraham and G. Snider. *PROC. (2000), on Conference International P 113-124. July Massachussetts, (Boston, 2000) (ASAP Processors and Architectures, Systems, Application-Specific)*. 2. Дунець Б.Р., Мельник А.О Високорівневе проектування НВІС з використанням параметризованої моделі спеціалізованої багатопроцесорної системи // "Комп'ютерні технології друкарства". – 2001. – №6. – С. 282–289. 3. Дунець Б.Р. Високорівневе проектування НВІС з використанням мови програмування LISP // Вісник ДУ "Львівська політехніка". – 1999. – №373. – С.118–122.