

ОПТИМИЗАЦИЯ БЫСТРОДЕЙСТВИЯ АРИФМЕТИКО-ЛОГИЧЕСКОГО УСТРОЙСТВА, РЕАЛИЗОВАННОГО НА ПЛИС

Жуковский А.Д., Халиуллина М.Н.

Национальный детский технопарк, г. Минск, Республика Беларусь

Ловшенко И.Ю.

В данной работе рассматриваются методы оптимизации арифметико-логических устройств (АЛУ) при их проектировании на базе программируемых логических интегральных схем (ПЛИС). Цель работы – выявление наиболее эффективных методов минимизации критического пути в АЛУ. На основе теоретического анализа процессорных архитектур предложены подходы к повышению тактовой частоты устройства. Исследовано влияние конвейеризации и использования специализированных аппаратных ресурсов ПЛИС (DSP-блоки (англ. Digital Signal Processing), встроенные цепи переноса) на общую производительность системы.

Основной задачей современных интегральных схем (микропроцессоров общего назначения, цифровых сигнальных процессоров и микроконтроллеров) является высокоскоростная обработка цифровых данных [1]. Центральным узлом вычислительного процесса традиционно выступает арифметико-логическое устройство (АЛУ).

Экспоненциальный рост объемов обрабатываемых данных требует сверхвысокой пропускной способности АЛУ, особенно при функционировании в составе систем на кристалле (англ. System on a Chip, SoC). При проектировании микропроцессоров разработчики сталкиваются со следующей проблемой: специфическая архитектура ПЛИС, базирующаяся на таблицах истинности (англ. Look-Up Tables, LUT), делает традиционные методы оптимизации, применяемые для заказных интегральных схем (англ. Application-Specific Integrated Circuit, ASIC), недостаточно эффективными [2].

Как отмечают авторы исследования [3], при использовании базовых АЛУ с последовательной организацией арифметического переноса, реализованных исключительно на базовой логике, время выполнения операции линейно возрастает с увеличением разрядности данных.

Производительность любого цифрового узла на ПЛИС определяется задержкой распространения сигнала по критическому пути – самому длинному маршруту от выхода одного регистра до входа следующего. Факторы, ограничивающие максимальную тактовую частоту, делятся на две категории: задержки логического уровня, которые определяются количеством последовательно соединенных LUT, которые сигнал должен пройти за один такт. В сложных арифметических операциях число уровней логики может достигать десятков; в современных технологических процессах (28 нм и ниже) задержки на межсоединениях внутри кристалла ПЛИС становятся доминирующим фактором, зачастую превышая задержки самих логических элементов.

Учитывая многофакторную природу этих задержек, решение проблемы минимизации критического пути не может сводиться к локальным изменениям в коде описания аппаратуры. Для достижения максимального быстродействия оптимизация должна проводиться комплексно на трех взаимосвязанных уровнях: алгоритмическом, структурном и аппаратном. Систематизация данных подходов и их ключевые направления представлены на рисунке 1.

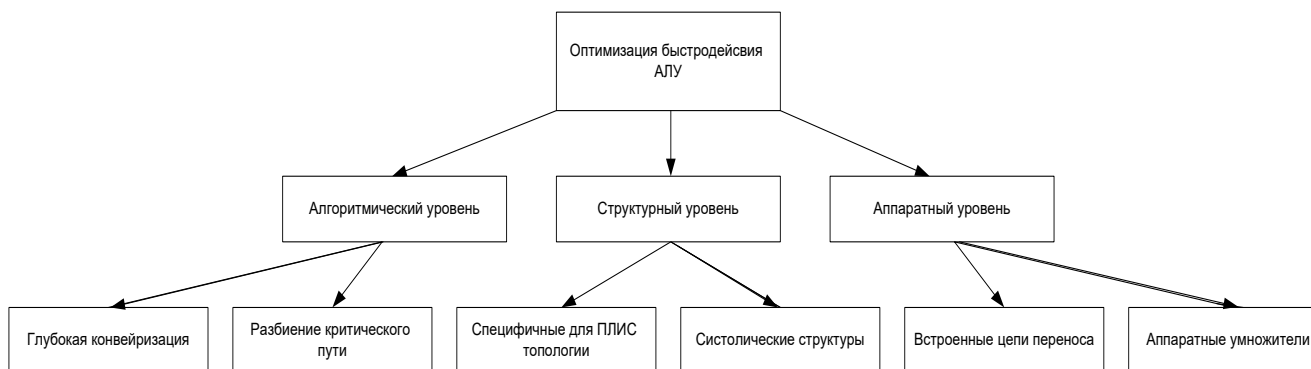


Рисунок 1 – методы оптимизации быстродействия АЛУ

Как следует из представленной классификации, каждый уровень абстракции предлагает собственный набор инструментов для повышения тактовой частоты. Однако наибольшую эффективность в условиях специфичной архитектуры ПЛИС демонстрируют решения, объединяющие структурные преобразования алгоритма с грамотным использованием заложенных производителем возможностей кристалла. Целесообразно рассматривать ключевые из этих методов, обеспечивающих наиболее существенный прирост производительности системы.

Для многоразрядных систем целесообразно использовать АЛУ с ускоренной организацией арифметического переноса, например, с параллельным (предвосхищающим) переносом [4, 5]. Однако при реализации на ПЛИС важно учитывать аппаратную специфику: синтез параллельного переноса на обычных LUT уступает по скорости встроенным решениям. Современные кристаллы имеют высокоскоростные выделенные аппаратные цепи переноса (Carry Chains, например, CARRY4/CARRY8 в архитектуре Xilinx). Переход от абстрактного логического описания сумматоров к структурам, которые САПР напрямую отображает на встроенные цепи переноса, дает значительный прирост быстродействия.

Архитектуры типа RISC исторически используют глубокую конвейеризации. Вставка промежуточных регистров (pipeline registers) в длинную комбинационную логику АЛУ позволяет разбить критический путь на несколько более коротких участков. Для повышения пропускной способности элементы АЛУ могут быть организованы в виде систолических структур, где логика коммутации и элементарных преобразований перемежается с регистрами [6]. Применение конвейеризации требует нахождения баланса: увеличение глубины конвейера повышает тактовую частоту, но увеличивает латентность операции.

Сравнительный анализ производительности вычислительных узлов показывает, что реализация ресурсоемких операций на базовой логике (LUT) неэффективна [7]. Для операций умножения, сдвига и накопления (MAC-операции), являющихся частью сложных АЛУ, необходимо использовать специализированные аппаратные DSP-блоки (например, DSP48E2 в Xilinx). Это жесткие IP-ядра (Hard IP), оптимизированные на уровне кремния, способные работать на предельных для кристалла частотах и освобождающие логику общего назначения.

Специфика ПЛИС требует отказа от прямого переноса ASIC-архитектур. Выявлено, что максимальное быстродействие арифметико-логического устройства на кристаллах ПЛИС достигается исключительно за счет синергии нескольких методов: корректного использования цепей ускоренного арифметического переноса (Carry Chains), глубокой алгоритмической конвейеризации и применения жестких аппаратных IP-блоков (DSP). Практическая значимость результатов заключается в возможности использования предложенных подходов для синтеза высокопроизводительных микропроцессорных ядер и вычислительных ускорителей для систем на кристалле.

При выборе метода оптимизации АЛУ разработчику необходимо учитывать баланс между занимаемой площадью кристалла и быстродействием.

В таблице 1 представлены обобщенные данные по изменению характеристик устройства при последовательном применении методов оптимизации (относительно базовой реализации).

Таблица 1 – Эффективность применения комплексной оптимизации

Этап оптимизации	Прирост максимальной тактовой частоты, %	Изменение площади, %	Примечание
Начальная реализация	0	0	Использование последовательного переноса, высокая задержка
Использование цепей переноса	+15–25	-5–10	Снижение нагрузки на LUT за счет выделенных аппаратных ресурсов
Внедрение конвейеризации	+40–65	+20–35	Рост частоты за счет сокращения критического пути; расход триггеров
Использование DSP-блоков	+80–120%	-40–60%	Максимальное быстродействие; высвобождение логики общего назначения

Список использованных источников:

- Zimmermann, R. *Binary adder architectures for cell-based VLSI and their synthesis* / R. Zimmermann. – Zurich: Swiss Federal Institute of Technology, 1997. – 110 p.
- Hauck, S. *Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation* / S. Hauck, A. DeHon. – Morgan Kaufmann, 2010. – 945 p.
- Якунин, А. Н. *Повышение быстродействия многоразрядного арифметико-логического устройства* / А. Н. Якунин, А. М. Сан, К. Вин. // *Известия высших учебных заведений. Электроника*. – 2021. – Т. 26, № 1. – С. 40–53.
- Моделирование микропроцессорных систем на базе программируемых логических интегральных схем с использованием Verilog HDL и САПР Quartus II : учеб. пособие по курсу «Микропроцессор. средства и системы» / Д. Н. Беклемишев, А. Н. Орлов, М. Г. Попов [и др.] ; под ред. А. Л. Переверзева. – М. : МИЭТ, 2014. – 100 с.
- Микропроцессорные средства и системы: курс лекций / Д. Н. Беклемишев, А. Н. Орлов, А. Л. Переверзев [и др.] ; Под ред. Ю.В. Савченко. – М. : МИЭТ, 2013. – 288 с.
- Федюнин, Р. Н. *Способ реализации аппаратного слоя VLIW-архитектуры на базе систолических структур* / Р. Н. Федюнин // *Известия высших учебных заведений. Поволжский регион. Технические науки*. – 2013. – № 2 (26). – С. 15–22.
- Underwood, K. D. *FPGAs vs. CPUs: trends in peak floating point performance* / K. D. Underwood // *Proceedings of the 2004 ACM/SIGDA 12th international symposium on Field programmable gate arrays*. – 2004. – P. 171–180.