

УДК 004.3:004.382.2

doi 10.25209/2079-3316-2025-16-1-45-59



О реализации QR-разложения на трехмерном систолическом массиве

Виктор Николаевич **Бабенко**^{1✉}, Артём Павлович **Невечеря**²

¹Кубанский государственный технологический университет, Краснодар, Россия

²Кубанский государственный университет, Краснодар, Россия

[✉]rnibvd@mail.ru

Аннотация. Интенсивные потоки данных, формирующихся систем линейных уравнений в режиме реального времени, а также системы линейных уравнений большой размерности обуславливают привлечение систолических массивов для их машинного решения. В представленном систолическом массиве, предназначенном для приведения матриц к треугольному виду, реализация ортогональных преобразований вращения может осуществляться как устройствами двумерного вращения вектора CORDIC, так и его модификациями. Для предложенного систолического массива даны описания его конфигурации, функционирования и технических характеристик, а также структуры входного и выходного потока данных.

Ключевые слова и фразы: систолический массив, функциональное устройство, продолжительность такта, маршрутизация и расписание движения данных, устройство вращения двумерного вектора, прямой ход

Для цитирования: Бабенко В.Н., Невечеря А.П. *О реализации QR-разложения на трехмерном систолическом массиве* // Программные системы: теория и приложения. 2025. Т. 16. № 1(64). С. 45–59. https://psta.psiras.ru/read/psta2025_1_45-59.pdf

Введение

Непрерывно возрастающие объемы вычислений обуславливают необходимость повышения производительности вычислительных систем. Одно из направлений в этой деятельности – разработка и применение систолических массивов (СМ). Вопросам разработки СМ различного назначения посвящено немало работ, так как они являются одним из наиболее эффективных средств решения указанной проблемы [1–8].

Одной из важнейших задач вычислительной линейной алгебры является решение систем линейных алгебраических уравнений (СЛАУ). Пусть матрица системы

$$Ax = f$$

невырождена, и N – ее порядок. Для осуществления QR-разложения матрицы A мы будем использовать преобразования (матрицы) вращения.

Введем обозначения $A^{(0,0)} = A$, $Q_{i,j}^*$ – преобразование вращения, аннулирующее i -ю компоненту j -ого столбца матрицы.

Для $j = \overline{1, N-1}$, $i = \overline{j, N}$ вычисляем:

$$x = a_j^{(j-1, i-1)}, \quad c = \frac{x_j}{\sqrt{x_j^2 + x_i^2}}, \quad s = \frac{x_i}{\sqrt{x_j^2 + x_i^2}},$$

$$\text{для } k = \overline{j, N}: \quad x = a_k^{(j-1, i-1)}, \quad y = S_{j,i}^* x, \quad a_k^{(j,i)} = y.$$

По завершении этой процедуры получим верхнюю треугольную матрицу $R = A^{(N-1, N)}$, связанную с исходной соотношением $B = S^* A$, где

$$S = S_{1,2} \dots S_{1,N} S_{2,3} \dots S_{2,N} \dots S_{N-2,N-1} S_{N-2,N} S_{N-1,N}.$$

Следует сказать, что в литературе для представления матрицы A в виде произведения $A = QR$, где Q и R – ортогональная и верхняя треугольная матрицы соответственно, широко применяется термин QR-разложение. Следуя этой традиции, произведение $Q^* A$ мы будем называть QR-упрощением матрицы A .

Представленный способ QR-разложения матрицы системы уравнений может быть эффективно реализован аппаратно в системах реального времени, формирующих плотные потоки данных для решения СЛАУ с плотными матрицами небольшой размерности.

Второй областью применения QR-разложения матрицы, очевидно, являются задачи решения СЛАУ большой размерности с разреженными матрицами. Для решения таких систем применяются итерационные методы. Пусть x^0 – произвольный вектор, $r^0 = f - Ax^0$ и

$$x^{n+1} = x^n + y^n, \quad n = 0, 1, 2, \dots$$

В проекционных методах вычисление поправки y^n сводится к последовательному решению трех систем линейных уравнений

$$V_n w^n \cong r^n, \quad H_n z^n \cong w^n, \quad V_n^* y^n \cong z^n,$$

где V_n – псевдоортогональная матрица, $r^n = f - Ax^n$ и H_n – некоторая плотная невырожденная матрица порядка m , причем m невелико.

Из трех выше представленных систем наиболее трудоемким является решение второй системы. Системы линейных уравнений $H_n z^n \cong w^n$, $n = 0, 1, 2, \dots$, обозначенного потока могут быть решены с помощью QR-разложения матрицы H_n , реализованного аппаратурно на специально разработанном вычислительном устройстве (систолическая матрица Джентльмена–Кунга [9]).

Обоснование заявленного утверждения упрощенно раскроем следующим образом. Для реализации преобразования вращения вектора разработано специальное вычислительное устройство CORDIC, которое производится в двух вариантах: $R1$, аннулирующее вторую компоненту входного вектора, и $R2$, работающее в режиме «эхо».

На вход $R1$ поступают компоненты x и y двумерного вектора $\begin{bmatrix} x \\ y \end{bmatrix}$. На выходе получают вектор $\begin{bmatrix} z \\ 0 \end{bmatrix}$, где $z = \text{sign}(x)\sqrt{x^2 + y^2}$, и параметры преобразования вращения c и s .

На вход $R2$ поступают компоненты u и v двумерного вектора $\begin{bmatrix} u \\ v \end{bmatrix}$ и параметры преобразования вращения c и s . Соответствующего на выходе получают вектор $\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$ и параметры преобразования вращения c и s .

Устройство CORDIC осуществляет вычисление компонент вектора за $2m$ тактов, где m – количество разрядов, отводимых под мантиссу машинных чисел. Устройства CORDIC получили применение в систолической структуре (матрице Джентльмена–Кунга), применяемой для приведения матрицы к верхнему треугольному виду. Эти устройства в конвейерном варианте способны обрабатывать потоки данных высокой плотности, но размерность систолической структуры Джентльмена–Кунга (она равна 2) препятствует реализации этой способности указанного устройства. Выходом из этого положения может служить переход к применению трехмерных систолических структур.

Следует сказать, что недостатком метода CORDIC является не слишком высокая скорость сходимости (для достижения приемлемой

точности требуется $2m$ итераций). Соответственно этому задержка вычисленного результата на выходе устройства CORDIC составит $2m$ тактов. Одним из авторов этой статьи разработаны модификация метода CORDIC, позволившая сократить число итераций до m , и соответствующее ей устройство, обеспечивающее возможность получить вычисленный результат за m тактов [10–13]).

Выполняя эту работу, мы преследовали цель обеспечить обработку потоков данных высокой плотности при осуществлении решения СЛАУ. Достижение этой цели мы видели в решении задачи разработки систолических структур, использующих в качестве функциональных устройств (ФУ) CORDIC-устройства конвейерного типа или их указанную выше модификацию [10–12].

1. Основные положения

В систолических массивах в состав ФУ входят функциональные элементы (ФЭ), которые под воздействием входных сигналов и синхронизирующих импульсов, называемых тактовыми, переходят из предыдущего состояния в последующее.

Пусть T_m и T_{m+1} – моменты поступления на ФЭ m -го и $m+1$ -го тактовых импульсов. Промежуток времени $\Delta T = T_{m+1} - T_m$ называют m -м тактом.

Символами $I_{ФЭ}$ и $O_{ФЭ}$ обозначают вход и выход ФЭ. Обозначение $I_{ФЭ}(m) = a$ означает, что на вход ФЭ на m -ом такте (обычно в начале промежутка ΔT) поступает сигнал (число) a .

Под воздействием сигнала a и m -го тактового импульса ФЭ с течением времени Δt переходит в новое устойчивое состояние, которое наблюдают на его выходе. При этом пишут $O_{ФЭ}(m) = b$, где b – сигнал (число) на выходе ФЭ на m -ом такте, соответственно Δt называют временем срабатывания ФЭ. Выполнение неравенства $\Delta t \leq \Delta T$ называют условием срабатывания ФЭ.

Переходя к формальному изложению, последнее мы можем записать так.

ЛЕММА 1.1 (о срабатывании ФЭ). Пусть $I_{ФЭ}(m) = a$. Если $\Delta t \leq \Delta T$, то $O_{ФЭ}(m) = b$.

Выражение $I_{ФЭ_2} \Leftarrow O_{ФЭ_1}$ ($O_{ФЭ_1} \Rightarrow I_{ФЭ_2}$) означает, что выход ФЭ₁ соединен с входом ФЭ₂.

ОПРЕДЕЛЕНИЕ 1.1. Если на m -том такте на выходе ФЭ появляется результат b (т.е. $O_{\Phi\mathcal{E}}(m) = b$), то его мы будем называть тактом, порождающим результат b при отображении ФЭ (обозначение $m = \text{ind}_{\Phi\mathcal{E}}(b)$).

ЛЕММА 1.2 (о локальном взаимодействии ФЭ). Пусть $I_{\Phi\mathcal{E}_2} \leq O_{\Phi\mathcal{E}_1}$. Если $I_{\Phi\mathcal{E}_1}(m) = a$ и $O_{\Phi\mathcal{E}_1}(m) = b$, то $I_{\Phi\mathcal{E}_2}(m+1) = b$.

ОПРЕДЕЛЕНИЕ 1.2. Конвейером C мы будем называть функциональное устройство, представляющее собой последовательность ФЭ

$$C = \{F_h, \dots, F_1\} : I_{F_{\rho+1}} \leq O_{F_{\rho}}, \rho = \overline{1, h-1}.$$

Соответственно $I_C = I_{F_1}$ – вход, $O_C = O_{F_h}$ – выход, и h – длина конвейера C .

Входящие в состав конвейера (СМ) функциональные элементы в общем случае имеют различное время срабатывания. Вопрос о выборе продолжительности такта, общей для конвейера, разрешает следующая

ЛЕММА 1.3 (о необходимом условии функционирования конвейера). Пусть конвейерный вычислитель C с регулярным графом состоит из n функциональных элементов, причем $\Delta t_i, i = \overline{1, h}$ – время срабатывания i -го элемента, и справедливы соотношения: $a_1 = F_1(a)$, $a_2 = F_2(a_1)$, \dots , $a_{h-1} = F_{h-1}(a_{h-2})$, $b = F_h(a_{h-1})$. Если $I_C(m) = a$, и $\Delta t \leq \Delta T$, где $\Delta t \geq \max\{\Delta t_1, \dots, \Delta t_h\}$, то $O_C(m+h-1) = a_h = b$.

Действуя по индукции, очевидно, применяя при этом леммы 1.1 и 1.2, мы докажем утверждение последней леммы.

Если $I_R(m) = a$ и $O_R(m) = a$, то функциональный элемент R называют элементом задержки, а конвейер LD , представляющий собой последовательность элементов задержки $R = \{R_h, \dots, R_1\} : I_{R_{\rho+1}} \leq O_{R_{\rho}}, \rho = \overline{1, h-1}$, $I_{LD} = I_{R_1}$, $O_{LD} = O_{R_h}$, называют линией задержки, соответственно I_{LD} – входом, O_{LD} – выходом и h – длиной линии задержки.

2. QR-упрощение матрицы на трехмерном СМ

Поскольку назначением предлагаемого ниже систолического массива прямого хода (СМПХ) является приведение матрицы системы $Ax = y$ к треугольному виду (этап 1) прежде всего для реализации последующего обратного хода (этап 2), и при осуществлении этапа 1 происходит трансформация вектора y , как и всех вектор-столбцов матрицы A , мы

сформируем расширенную матрицу $[A \ y]$. Эту матрицу, как и ранее, мы будем обозначать символом A , а вектор y считать $(N+1)$ -м столбцом расширенной матрицы A .

2.1. Конфигурация СМПХ

Для реализации QR-упрощения мы применим трехмерную матрицу ФУ R , состав элементов которой, описывается соотношениями:

$$R_{ijk} = \begin{cases} LD & \forall j, \text{ если } i = k, \\ R1 & \forall i, \text{ если } j = k, \\ R2 & \text{в остальных случаях,} \end{cases}$$

где

$$k = \overline{1, N-1}, \quad i = \overline{k, N}, \quad j = \overline{k, N+1}.$$

В целях упрощения (единообразия) дальнейшего изложения мы будем придерживаться следующих соглашений:

$$I_{R_{i,j,k}} = I2_{R_{i,j,k}}, \quad O_{R_{i,j,k}} = O1_{R_{i,j,k}}, \quad k = \overline{1, N-1}, \quad i = k, \\ j = \overline{k, N+1}.$$

Функциональные устройства связаны между собой следующим образом.

(1) Одноуровневые (горизонтальные) связи:

$$O3_{R_{i,j,k}^\rho} \Rightarrow I3_{R_{i,j+1,k}^\rho}, \quad k = \overline{1, N-1}, \quad i = \overline{k+1, N}, \quad j = \overline{k, N}, \\ \rho = \overline{1, m}, \quad (\text{направление} - \text{вправо});$$

$$O1_{R_{i,j,k}} \Rightarrow I1_{R_{i+1,j,k}}, \quad k = \overline{1, N-1}, \quad i = \overline{k+1, N-1}, \\ j = \overline{k, N+1}, \quad (\text{направление} - \text{вперед («вниз»)).$$

(2) Межуровневые связи:

$$O2_{R_{i,j,k}} \Rightarrow I2_{R_{i,j,k+1}}, \quad k = \overline{1, N-1}, \quad i = \overline{k+1, N}, \\ j = \overline{k+1, N+1}, \quad (\text{направление} - \text{вниз}).$$

Входы СМПХ:

$$I2_{R_{1,j,k}}, \quad i = \overline{1, N}, \quad j = \overline{1, N+1}.$$

Выходы СМПХ:

$$O1_{R_{N,j,k}}, \quad k = \overline{1, N-1}, \quad j = \overline{k, N+1},$$

$$O2_{R_{N,N,N-1}}, \quad O2_{R_{N,N+1,N-1}}.$$

2.2. Основные свойства СМПХ (маршрутизация и расписание движения данных)

Пусть требуется осуществить QR-упрощение матриц

$$A^1, \dots, A^l, \dots.$$

Описание свойств СМПХ предполагает применение переменных:

t, \bar{t} – номера тактов,

$\tau, \bar{\tau}$ – номера упрощаемых матриц (решаемых систем уравнений).

Для описания свойств (технических характеристик) СМПХ, представляющихся нам основными, определим следующие величины:

$\Delta \bar{T}$ – продолжительность порождения упрощенной матрицы, определяющаяся по формуле $\Delta \bar{T} = k_e - k_b + 1$, где k_b – номер такта, на котором на вход СМПХ поступает первый элемент упрощаемой матрицы, k_e – номер такта, на котором порождается последний элемент упрощенной матрицы;

P – производительность, определяемая как количество элементов упрощаемых матриц систем уравнений, появляющихся на выходе (выходах) СМПХ одновременно.

ВЫСКАЗЫВАНИЕ 2.1. Пусть

(i) продолжительность ΔT такта на СМПХ удовлетворяет неравенству

$$\Delta T \geq \Delta t_+,$$

где Δt_+ – продолжительность времени срабатывания сумматора.

(ii) для всех $l = 1, 2, \dots, i = \overline{1, N-1}, j = \overline{1, N+1}$ выполняются соотношения

$$(1) \quad I2_{R_{i,j,1}}(t) = a_{i,j}^{\tau}, \quad \text{где}$$

$$t = h(i-1) + l,$$

$$\tau = t - h(i-1) - (j-1),$$

причем соотношения (1) определены для всех $l, i, j : \tau > 0$.

Тогда

1) для всех $k = \overline{1, N-1}$

$$O1_{R_{N,j,k}}(\bar{t}) = b_{k,j}^{\bar{t}},$$

$$O2_{R_{N,N+1,N-1}}(h(N + (N-2) + (l-1))) = b_{N,N}^l,$$

$$O2_{R_{N,N+1,N-1}}(h(N + (N-2)) + l) = b_{N,N+1}^{l+1},$$

$$\begin{aligned} \bar{t} &= h(N + (k-1)) + (l-1), \\ \text{где } \bar{\tau} &= \bar{t} - h(N + (k-1)) - (j-1) + 1; \end{aligned}$$

2) продолжительность процедуры QR-упрощения матрицы $\Delta\bar{T}$ равна $2hN + N + 1$ тактам;

3) производительность СМПХ P равна $(N+3)N/2$ слов за такт.

ЗАМЕЧАНИЕ 2.1. Значения величин пунктов 2) и 3) высказывания 2.1 получены применением леммы 1.3.

Заключение

В представленном СМПХ за счет увеличения размерности СМ до трех достигнута максимально возможная плотность потоков движения данных. Она ограничена продолжительностью выполнения операции сложения Δt_+ . Сопоставляя упомянутую во введении матрицу Джентльмена–Кунга с СМПХ, мы видим, что производительность разработанного СМПХ в $n \times 2h$ выше, чем у СМ Джентльмена–Кунга. В n раз выше за счет увеличения размерности массива до трех, и в $2h$ раз – за счет появившейся при этом возможности увеличить плотность потока данных в $2h$ раз, где

$$h = \begin{cases} m & \text{для устройства CORDIC,} \\ m/2 & \text{для модифицированного CORDICa,} \end{cases}$$

m – количество двоичных разрядов, отводимых под мантиссу числа.

Проводя обзор публикаций последних лет по реализации QR-разложения на трехмерных систолических массивах с использованием CORDIC-устройств, мы обнаружили следующее положение дел.

В работе [14] предложена архитектура двумерного систолического массива, который может быть использован в сверхбольших интегральных схемах (СБИС) для аппаратной реализации алгоритма QR-разложения комплексных матриц. Предложенная архитектура является эффективной для реализации в СБИС, так как использует только алгоритм CORDIC и, следовательно, не использует операций умножения. Рассматриваемый систолический массив основан на методе треугольных комплексных

вращений и позволяет получить значительный выигрыш в производительности по сравнению с традиционно используемым методом комплексных вращений Гивенса.

В статье [15] исследуется специфическое исключение элементов для решения линейной алгебраической системы вида $Az = b$, где матрица A задана суммой m парных произведений трёхдиагональных (нижней и верхней) матриц порядка n . Предложен алгоритм, требующий лишь $(3m - 2)n^2$ умножений и такого же числа сложений. Для симметричной матрицы A предложен алгоритм, требующий лишь $(2m - 1)n^2$ умножений и такого же числа сложений. Алгоритмы обладают регулярной структурой и допускают параллельную реализацию за $O(nm)$ шагов.

В работах иностранных авторов представлены результаты по практической реализации QR-разложения на двумерных систолических массивах.

В статье [16] дано описание двумерного систолического массива, реализованного на ПЛИС Xilinx Virtex5 и предназначенного для реализации QR-разложения с использованием алгоритма вращения Гивенса.

В работе [17] представлена улучшенная архитектура систолического массива для реализации QR-разложения на основе метода вращения Гивенса (GR) для действительной матрицы 4×4 . Алгоритм цифрового компьютера вращения координат (CORDIC) принят и модифицирован для ускорения и упрощения процесса GR.

В статье [18] представлен улучшенный аппаратный дизайн с фиксированной точкой QR-разложения, специально оптимизированный для ПЛИС Xilinx. Алгоритм вращения Гивенса реализован с использованием свернутого систолического массива и алгоритма CORDIC.

В статье [19] представлена параллельная архитектура систолического массива QR-разложения на основе алгоритма вращения Гивенса на ПЛИС. Предлагаемая архитектура использует прямое отображение 21 процессорных единиц с фиксированной точкой на основе CORDIC, которые могут вычислять QR-разложение для действительной матрицы 4×4 .

По трёхмерным систолическим массивам мы обнаружили только три работы

В [20] предложена архитектура трёхмерного систолического вычислителя дискретного преобразования Фурье на основе кронекеровского произведения матриц.

Авторы в статье [21] сообщают, что ими обнаружено следующее. Систолический массив как уровень программной виртуализации может

привести к чрезвычайно масштабируемой парадигме выполнения. Чтобы продемонстрировать эту масштабируемость, они спроектировали и реализовали виртуальный трехмерный систолический массив для реализации QR-разложения. На машине Cray-XT5 продемонстрировано, что виртуальный систолический массив может обеспечить превосходную параллельную производительность.

В работе [22] рассмотрена реализация HLS архитектуры трехмерного систолического массива для умножения матриц, которая нацелена на определенные характеристики ПЛИС Intel Stratix 10, чтобы создавать проекты, которые достигают высокой пропускной способности с плавающей точкой.

Представленный анализ публикаций последних лет позволяет сделать вывод, что полученные в нашей работе результаты являются новыми.

Подведем итоги проделанной работы. Для приведения матрицы СЛАУ к верхнему треугольному виду разработан трехмерный СМ, ориентированный на применение в системах реального времени с жесткими требованиями ко времени реакции, а также в вычислительных системах, осуществляющих решение СЛАУ большой размерности. Выполненная работа имеет очевидное продолжение – проведение проектных работ для последующего изготовления разработанного СМ в «железе».

Список использованных источников

- [1] Елфимова Л. Д. *Быстрые алгоритмы для базовой операции клеточных методов линейной алгебры* // Кибернетика и системный анализ.– 2015.– Т. 51.– № 6.– С. 35–45.  [↑46](#)
- [2] Горюшкина А. Е., Семенов С. Г. *Применение систолической системы на основе метода моментов для усовершенствования дискретного преобразования Хартли* // Збірник наукових праць Харківського університету Повітряних Сил.– 2016.– № 1(46).– С. 89–91.  [↑46](#)
- [3] Azimian A., Dehkordi A. K., Tehrani M. *A novel systolic array architecture for matrix multiplication circuit design using carbon nanotube technology* // International Journal of Computer Applications.– 2017.– Vol. 172.– No. 6.– Pp. 1–4.  [↑46](#)
- [4] Causapruno G., Riente F., Turvani G., Vacca M., Massimo R. R., Zamboni M., Graziano M. *Reconfigurable systolic array: from architecture to physical design for NML* // IEEE Transactions on Very Large Scale Integration (VLSI) Systems.– 2016.– Vol. 24.– No. 11.– Pp. 3208–3217.   [↑46](#)
- [5] Randjelovic B. M., Milovanovic E. I., Milovanovic I. Z. *Systolic algorithms for matrix multiplication on space optimal one-dimensional systolic arrays* // Facta Universitatis (Nis). Series: Mathematics and Informatics.– 2014.– Vol. 29.– No. 3.– Pp. 243–259.  [↑46](#)

- [6] Abdollahi M. M., Tehrani M. *Designing a novel reversible systolic array using QCA* // Italian Journal of Science & Engineering.– 2017.– Vol. 1.– No. 3.– Pp. 158–166. ^{↑46}
- [7] Воеводин В. В. *Математические модели и методы в параллельных процессах.*– М.: Наука.– 1986.– 296 с. ^{↑46}
- [8] Бабенко В. Н., Ивановский О. Я., Тимонов Д. А. *Систематическая структура обратного хода метода Гаусса и ее основные свойства* // Экологический вестник научных центров Черноморского экономического сотрудничества.– 2018.– Т. 15.– № 1.– С. 8–14. ^{↑46}
- [9] С. Гун, Х. Уайтхаус, Т. Кайлат (ред.) *Сверхбольшие интегральные схемы и современная обработка сигналов*, пер. с англ. под ред. В. А. Лексаченко.– М.: Радио и связь.– 1989.– ISBN 5-256-00179-5.– 472 с. ^{↑47}
- [10] Бабенко В. Н. *Устройство вращения вектора*, Пат. 2475830 Российская Федерация, МПК G 06 F 17/16, № 2010134147/08, заявл. 13.08.10, опубл. 20.02.13, Бюл. № 5.– 11 с. ^{↑48}
- [11] Бабенко В. Н. *Устройство нормировки вектора*, Пат. 2449354 Российская Федерация, МПК G 06 F 17/16, № 2010134146/08, заявл. 13.08.10, опубл. 27.04.12, Бюл. № 12.– 7 с. ^{↑48}
- [12] Бабенко В. Н. *Алгоритм инверсии делителя* // Экологический вестник научных центров Черноморского экономического сотрудничества.– 2013.– Т. 10.– № 4–1.– С. 19–25. ^{↑48}
- [13] Бабенко В. Н. *Аспекты машинного решения систем линейных уравнений*, монография.– Краснодар: Экоинвест.– 2020.– ISBN 978-5-94215-531-5.– 307 с. ^{↑48}
- [14] Мальцев А. А., Пестрецов В. А., Хоряев А. В., Масленников Р. О. *Метод QR-разложения комплексной матрицы с помощью треугольного систолического массива* // Сборник научных трудов «Проблемы разработки перспективных микроэлектронных систем» (9–12 октября 2006 г.), ред. А. Л. Стемпковский, М.: ИПИМ РАН.– 2006.– С. 91–96. ^{↑52}
- [15] Тыртышников Е. Е. *Параллельные методы для обобщенно-тёплицевых систем* // Журнал вычислительной математики и математической физики.– 1996.– Т. 36.– № 6.– С. 5–19. ^{↑53}
- [16] Xiaojun W., Miriam L. *A truly two-dimensional systolic array FPGA implementation of QR decomposition* // ACM Transactions on Embedded Computing Systems.– 2009.– Vol. 9.– No. 1.– id. 3.– 17 pp. ^{↑53}
- [17] Jianfeng Z., Paul C., Hengzhu L. *CORDIC-based enhanced systolic array architecture for QR decomposition* // ACM Transactions on Embedded Computing Systems.– 2015.– Vol. 9.– No. 2.– id. 9.– 22 pp. ^{↑53}
- [18] Semih A., Sufeng N., Jafar S. *FPGA implementation of fast QR decomposition based on givens rotation* // 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS) (05–08 August 2012, Boise, ID, USA).– IEEE.– 2012.– ISBN 9781467325264.– Pp. 470–473. ^{↑53}
- [19] Dongdong C., Mihai S. *Fixed-point CORDIC-based QR decomposition by givens rotations on FPGA* // 2011 International Conference on Reconfigurable Computing and FPGAs (30 November 2011 – 02 December 2011, Cancun, Mexico).– IEEE.– 2012.– ISBN 978-0-7695-4551-6.– Pp. 327–332. ^{↑53}

- [20] Григорьев В. Р. *Методы параллельной цифровой обработки информации в трехмерных оптических интегральных схемах*, Дис. ... канд. техн. наук : 05.13.17.– М.– 2005.– 231 с.  [↑53](#)
- [21] Yamazaki I., Kurzak J., Luszczek P., Dongarra J. *Design and implementation of a large scale tree-based QR decomposition using a 3D virtual systolic array and a lightweight runtime* // *Parallel Processing Letters*.– 2014.– Vol. **24**.– No. 04.– id. 1442004.  [↑53](#)
- [22] Paolo G., Christian P. *High level synthesis implementation of a three-dimensional systolic array architecture for matrix multiplications on Intel Stratix 10 FPGAs*.– 2021.– 9 pp. arXiv  2110.11521 [cs.AR]  [↑54](#)

Поступила в редакцию 28.11.2024;
 одобрена после рецензирования 04.12.2024;
 принята к публикации 18.02.2025;
 опубликована онлайн 24.02.2025.

Рекомендовал к публикации

к.т.н. С. А. Амелькин

Информация об авторах:



Виктор Николаевич Бабенко

к.ф.-м.н., доц. кафедры общей математики Кубанского государственного технологического университета. Научные интересы: численные методы.

 0009-0002-8341-0170
 e-mail: nibvd@mail.ru



Артём Павлович Невечеря

ст. преподаватель кафедры математических и компьютерных методов Кубанского государственного университета. Научные интересы: математическое моделирование динамики незамкнутых сложных систем, экстремальные задачи, системы искусственного интеллекта.

 0000-0001-6736-4691
 e-mail: artiom1989@mail.ru

Вклад авторов: *В. Н. Бабенко* – 75% (идея, методология, написание черновой версии); *А. П. Невечеря* – 25% (валидация, доработка и редактирование).

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*

UDC 004.3:004.382.2

 10.25209/2079-3316-2025-16-1-45-59

On the implementation of QR-decomposition on a three-dimensional systolic array

Viktor Nikolaevich **Babenko**¹, Artyom Pavlovich **Nevecherya**²

¹ Kuban State Technological University, Krasnodar, Russia

² Kuban State University, Krasnodar, Russia

 rnibvd@mail.ru

Abstract. Intensive data flows, formed systems of linear equations in real time, as well as systems of linear equations of large dimensionality cause the involvement of systolic arrays for their machine solution. In the presented systolic array, designed to reduce matrices to triangular form, the realization of orthogonal rotation transformations can be carried out both by two-dimensional vector rotation devices CORDIC, and its modifications. For the proposed systolic array, descriptions of its configuration, operation and technical characteristics, as well as the structure of input and output data flow are given. (*In Russian*).

Key words and phrases: systolic array, functional unit, cycle duration, data routing and scheduling, two-dimensional vector rotation unit, forward stroke

2020 *Mathematics Subject Classification:* 68M20; 37E45, 68M14

For citation: Viktor N. Babenko, Artyom P. Nevecherya. *On the implementation of QR-decomposition on a three-dimensional systolic array.* Program Systems: Theory and Applications, 2025, **16**:1(64), pp. 45–59. (*In Russ.*). https://psta.pspiras.ru/read/psta2025_1_45-59.pdf

References

- [1] L. D. Elfimova. “Fast algorithms for the basic operation of cellular linear algebra methods”, *Kibernetika i sistemnyj analiz*, **51:6** (2015), pp. 35–45 (in Russian).
- [2] A. E. Goryushkina, S. G. Semenov. “Application of the systolic system based on the method of moments to improve the discrete Hartley transform”, *Zbirnik naukovix prac' Xarkivs'kogo universitetu Povitryanix Sil*, 2016, no. 1(46), pp. 89–91 (in Russian).
- [3] A. Azimian, A. K. Dehkordi, M. Tehrani. “A novel systolic array architecture for matrix multiplication circuit design using carbon nanotube technology”, *International Journal of Computer Applications*, **172:6** (2017), pp. 1–4. 
- [4] G. Causaprano, F. Riente, G. Turvani, M. Vacca, R. R. Massimo, M. Zamboni, M. Graziano. “Reconfigurable systolic array: from architecture to physical design for NML”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, **24:11** (2016), pp. 3208–3217.  
- [5] B. M. Randjelovic, E. I. Milovanovic, I. Z. Milovanovic. “Systolic algorithms for matrix multiplication on space optimal one-dimensional systolic arrays”, *Facta Universitatis (Nis). Series: Mathematics and Informatics*, **29:3** (2014), pp. 243–259. 
- [6] M. M. Abdollahi, M. Tehrani. “Designing a novel reversible systolic array using QCA”, *Italian Journal of Science & Engineering*, **1:3** (2017), pp. 158–166. 
- [7] V. V. Voevodin. *Mathematical Models and Methods in Parallel Processes*, Nauka, M., 1986, 296 pp. (in Russian).
- [8] V. N. Babenko, O. Ya. Ivanovskij, D. A. Timonov. “Systolic structure of the Gauss method return stroke and its main properties”, *Ekologicheskij vestnik nauchnyx centrov Chernomorskogo ekonomicheskogo sotrudnichestva*, **15:1** (2018), pp. 8–14 (in Russian). 
- [9] S. Y. Kung, H. J. Whitehouse. *VLSI and Modern Signal Processing*, ed. T. Kailath, Prentice Hall, Englewood Cliffs, NJ, 1985, ISBN 013942699X, xiii+481 pp.
- [10] V. N. Babenko. *Vector rotation device*, Pat. 2475830 Rossijskaya Federaciya, MPK G 06 F 17/16, No 2010134147/08, zayavl. 13.08.10, opubl. 20.02.13, Byul. No 5, 11 pp. (in Russian).
- [11] V. N. Babenko. *Vector normalization device*, Pat. 2449354 Rossijskaya Federaciya, MPK G 06 F 17/16, No 2010134146/08, zayavl. 13.08.10, opubl. 27.04.12, Byul. No 12, 7 pp. (in Russian).
- [12] V. N. Babenko. “Divisor inversion algorithm”, *Ekologicheskij vestnik nauchnyx centrov Chernomorskogo ekonomicheskogo sotrudnichestva*, **10:4-1** (2013), pp. 19–25 (in Russian).
- [13] V. N. Babenko. *Aspects of Machine Solution of Systems of Linear Equations*, monograph, Ekoinvest, Krasnodar, 2020, ISBN 978-5-94215-531-5, 307 pp. (in Russian).

- [14] A. A. Mal'cev, V. A. Pestrecov, A. V. Xoryaev, R. O. Maslennikov. "QR decomposition method of complex matrix using triangular systolic array", *Sbornik nauchnyx trudov "Problemy razrabotki perspektivnyx mikroelektronnyx sistem"* (9–12 oktyabrya 2006 g.), ed. A. L. Stempkovskij, IPPM RAN, M., 2006, pp. 91–96 (in Russian). [URL](#)
- [15] E. E. Tyrtysnikov. "Parallel methods for generalized Toeplitz systems", *Comput. Math. Math. Phys.*, **36**:6 (1996), pp. 691–702.
- [16] W. Xiaojun, L. Miriam. "A truly two-dimensional systolic array FPGA implementation of QR decomposition", *ACM Transactions on Embedded Computing Systems*, **9**:1 (2009), id. 3, 17 pp. [doi](#)
- [17] Z. Jianfeng, C. Paul, L. Hengzhu. "CORDIC-based enhanced systolic array architecture for QR decomposition", *ACM Transactions on Embedded Computing Systems*, **9**:2 (2015), id. 9, 22 pp. [doi](#)
- [18] A. Semih, N. Sufeng, S. Jafar. "FPGA implementation of fast QR decomposition based on givens rotation", *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)* (05–08 August 2012, Boise, ID, USA), IEEE, 2012, ISBN 9781467325264, pp. 470–473. [doi](#) [URL](#)
- [19] C. Dongdong, S. Mihai. "Fixed-point CORDIC-based QR decomposition by givens rotations on FPGA", *2011 International Conference on Reconfigurable Computing and FPGAs* (30 November 2011 – 02 December 2011, Cancun, Mexico), IEEE, 2012, ISBN 978-0-7695-4551-6, pp. 327–332. [doi](#)
- [20] V. R. Grigor'ev. *Methods of parallel digital information processing in three-dimensional optical integrated circuits*, Dis. . . . kand. texn. nauk : 05.13.17, M., 2005, 231 pp. (in Russian). [URL](#)
- [21] I. Yamazaki, J. Kurzak, P. Luszczek, J. Dongarra. "Design and implementation of a large scale tree-based QR decomposition using a 3D virtual systolic array and a lightweight runtime", *Parallel Processing Letters*, **24**:04 (2014), id. 1442004. [doi](#)
- [22] G. Paolo, P. Christian. *High level synthesis implementation of a three-dimensional systolic array architecture for matrix multiplications on Intel Stratix 10 FPGAs*, 2021, 9 pp. arXiv:2110.11521~[cs.AR] [doi](#)