

Спецкурс кафедры «Вычислительной
математики»

**Параллельные алгоритмы
вычислительной алгебры**

Александр Калинин

Сергей Гололобов

Часть 5: Распараллеливание на компьютерах с распределенной памятью

- Linpack
- LAPACK
- DAG алгоритм

Blas

Basic Linear Algebra Subprograms

- BLAS Level 1 – операции с векторами (скалярное произведение вектор, умножение вектор на скалярную величину, нормы вектора и т.д.)
- BLAS Level 2 – матрично-векторные операции (умножение матрицы разных типов на вектор)
- BLAS Level 3 – операции с матрицами (перемножение прямоугольных матриц различных типов)

Опубликован в 1979 году

В свободном доступе на netlib.org

Содержится в оптимизированном виде в огромном количестве математических библиотек (Intel MKL, ACML, ATLAS, и тд)

Linpack

Linear Algebra Package

-Пакет для решения систем линейных уравнений и задачи о наименьших квадратах

Опубликован в конце 70-х годов Джеком Донгарра с коллективом

В свободном доступе на netlib.org

Впоследствии пакет стал основным замером производительности кластеров, в частности top500 определяются по модификации этого пакета.

Основная функциональность – разложение Холецкого, в симметричном случае представление матрицы

$$A = L \bullet L^T$$

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

$$=$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

$$\bullet$$

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

$$A_{11} = (L_{11})^2 \Rightarrow L_{11} = \sqrt{A_{11}}$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

$$=$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

$$\bullet$$

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

$$A_{1i} = L_{11} \cdot L_{1i} \Rightarrow L_{1i} = \frac{A_{1i}}{L_{11}}$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

$$=$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

$$\bullet$$

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

$$A_{22} = (L_{12})^2 + (L_{22})^2 \Rightarrow L_{22} = \sqrt{A_{22} - (L_{12})^2}$$

A_{11}	A_{12}	A_{13}	A_{14}		L_{11}	0	0	0		L_{11}	L_{12}	L_{13}	L_{14}
A_{12}	A_{22}	A_{23}	A_{24}		L_{12}	L_{22}	0	0		0	L_{22}	L_{23}	L_{24}
A_{13}	A_{23}	A_{33}	A_{34}	=	L_{13}	L_{23}	L_{33}	0	•	0	0	L_{33}	L_{34}
A_{14}	A_{24}	A_{34}	A_{44}		L_{14}	L_{24}	L_{34}	L_{44}		0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

$$A_{2i} = L_{1i} \cdot L_{12} + L_{2i} \cdot L_{22} \Rightarrow L_{2i} = \frac{A_{2i} - L_{1i} \cdot L_{12}}{L_{22}}$$

A_{11}	A_{12}	A_{13}	A_{14}	=	L_{11}	0	0	0	\bullet	L_{11}	L_{12}	L_{13}	L_{14}	
A_{12}	A_{22}	A_{23}	A_{24}		L_{12}	L_{22}	0	0		0	0	L_{22}	L_{23}	L_{24}
A_{13}	A_{23}	A_{33}	A_{34}		L_{13}	L_{23}	L_{33}	0		0	0	L_{33}	L_{34}	L_{34}
A_{14}	A_{24}	A_{34}	A_{44}		L_{14}	L_{24}	L_{34}	L_{44}		0	0	0	L_{44}	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

И так далее...

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

=

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

•

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

И так далее...

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

=

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

•

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

И так далее...

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

=

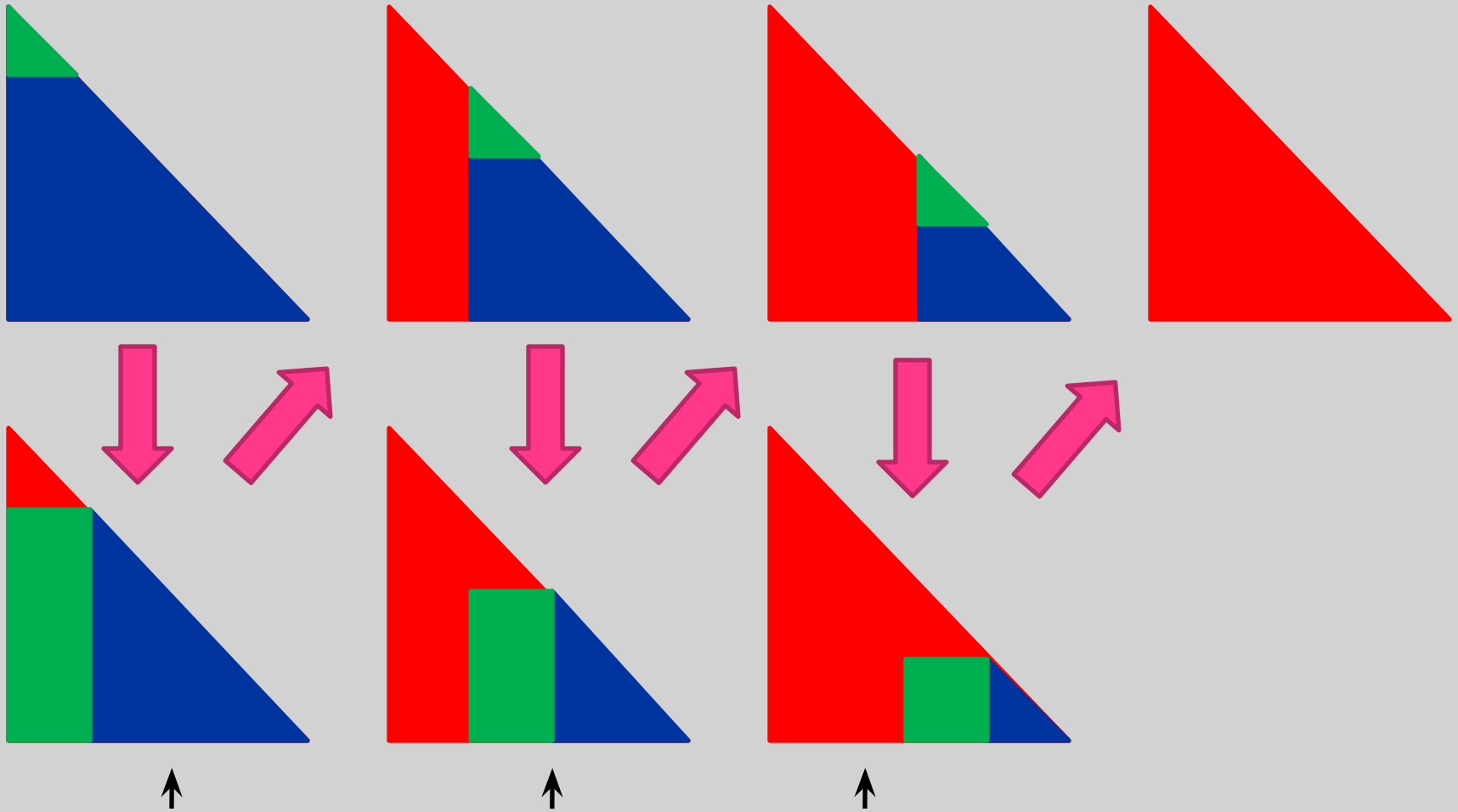
L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

•

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц



Может быть выполнено с помощью BLAS level1

Linpack

Плюсы:

- Достаточно оптимизировать BLAS level 1 для процессора, чтоб получить оптимизированный Linpack

Минусы:

- При увеличении кэша становится неэффективно умножать только строку на число – кэш значительно больше, есть возможность использовать его более разумно
- После каждого использования BLAS level 1 приходится вычислять корень из 1 вещественного числа – неэффективно для современных процессоров
- Blas level 1 не очень хорошо параллелизуется, появление многоядерных процессоров накладывает свои требования

LAPACK

Linear Algebra Package

-Пакет для решения систем линейных уравнений, поиска сингулярных значений матриц, задач о наименьших квадратах и многое другое...

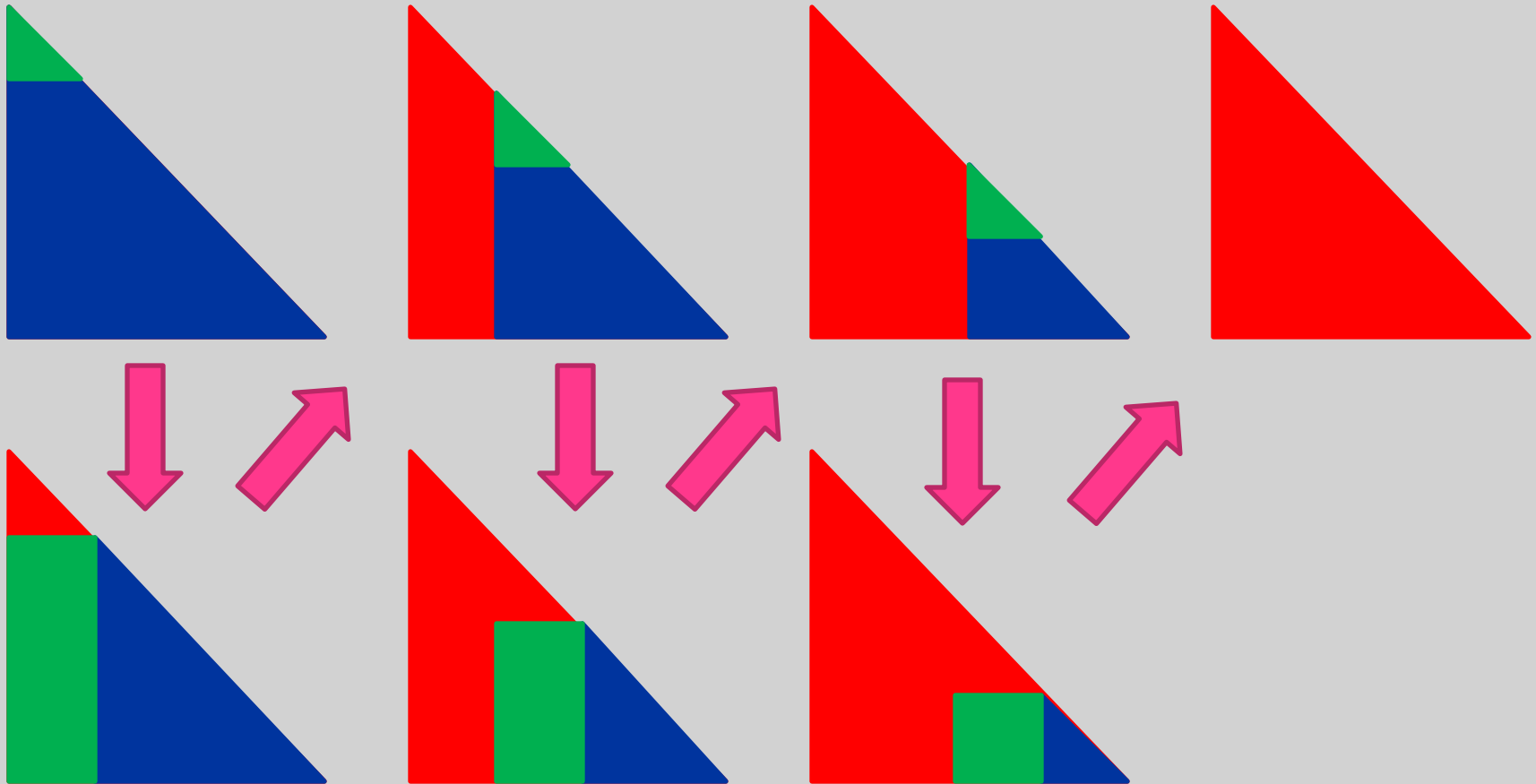
Опубликован в конце 1992 году Джеком Донгарра с коллективом

В свободном доступе на netlib.org

Содержится в оптимизированном виде в огромном количестве математических библиотек (Intel MKL, ACML, ATLAS, и тд)

Содержит параллельную версию разложения Холесского

LAPACK



Пусть каждый блок не один столбец, а группа. Тогда BLAS level 1 заменяется на BLAS level 3– за счет большего объема данных параллелизуется более эффективно чем level 1.

LAPACK

Плюсы:

- Достаточно оптимизировать BLAS level 3 для процессора, чтоб получить оптимизированное разложение Холецкого

Минусы:

- Не такая эффективная работа на процессорах с разным уровнем кэша – постоянно приходится перекачивать данные с уровня на уровень.
- Каждый эффективный вызов BLAS level 3 чередуется с неэффективным вызовом LL^T разложения для диагонального блока.
- При большом числе процессоров возникает ограничение на “шкалирование” вычисления группы столбцов – если группа большая, то время на вычисление диагонального блока становится существенным.

Решение проблемы с диагональным блоком

$$A = L \bullet L^T$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

 $=$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

 \bullet

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Решение проблемы с диагональным блоком

$$A = L \bullet L^T$$

$$A_{11} = (L_{11})^2 \Rightarrow L_{11} = \sqrt{A_{11}}$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

$$=$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

$$\bullet$$

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Решение проблемы с диагональным блоком

$$A = L \bullet L^T$$

$$A_{12} = L_{11} \cdot L_{12} \Rightarrow L_{12} = \frac{A_{12}}{L_{11}}$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

$$=$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

$$\bullet$$

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Решение проблемы с диагональным блоком

$$A = L \bullet L^T$$

$$A_{1i} = L_{11} \cdot L_{1i} \Rightarrow L_{1i} = \frac{A_{1i}}{L_{11}}$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

$$=$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

$$\bullet$$

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

Разложение Холецкого для симметричных матриц

$$A = L \bullet L^T$$

$$A_{22} = (L_{12})^2 + (L_{22})^2 \Rightarrow L_{22} = \sqrt{A_{22} - (L_{12})^2}$$

A_{11}	A_{12}	A_{13}	A_{14}		L_{11}	0	0	0		L_{11}	L_{12}	L_{13}	L_{14}
A_{12}	A_{22}	A_{23}	A_{24}		L_{12}	L_{22}	0	0		0	L_{22}	L_{23}	L_{24}
A_{13}	A_{23}	A_{33}	A_{34}		L_{13}	L_{23}	L_{33}	0		0	0	L_{33}	L_{34}
A_{14}	A_{24}	A_{34}	A_{44}		L_{14}	L_{24}	L_{34}	L_{44}		0	0	0	L_{44}

=

L_{11}	0	0	0		L_{11}	L_{12}	L_{13}	L_{14}
L_{12}	L_{22}	0	0		0	L_{22}	L_{23}	L_{24}
L_{13}	L_{23}	L_{33}	0		0	0	L_{33}	L_{34}
L_{14}	L_{24}	L_{34}	L_{44}		0	0	0	L_{44}

•

Красным обозначены известные значения
Синим неизвестные

Решение проблемы с диагональным блоком

$$A = L \bullet L^T$$

$$A_{1i} = L_{11} \cdot L_{1i} \Rightarrow L_{1i} = \frac{A_{1i}}{L_{11}}$$

$$A_{2i} = L_{1i} \cdot L_{12} + L_{2i} \cdot L_{22} \Rightarrow L_{2i} = \frac{A_{2i} - L_{1i} \cdot L_{12}}{L_{22}}$$

A_{11}	A_{12}	A_{13}	A_{14}
A_{12}	A_{22}	A_{23}	A_{24}
A_{13}	A_{23}	A_{33}	A_{34}
A_{14}	A_{24}	A_{34}	A_{44}

=

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

•

L_{11}	L_{12}	L_{13}	L_{14}
0	L_{22}	L_{23}	L_{24}
0	0	L_{33}	L_{34}
0	0	0	L_{44}

Красным обозначены известные значения
Синим неизвестные

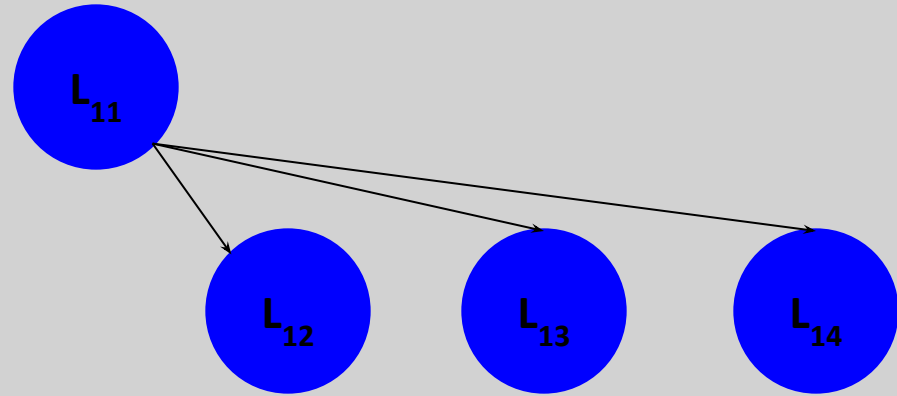
Дag подход (Directed acyclic graph)



L_{11}

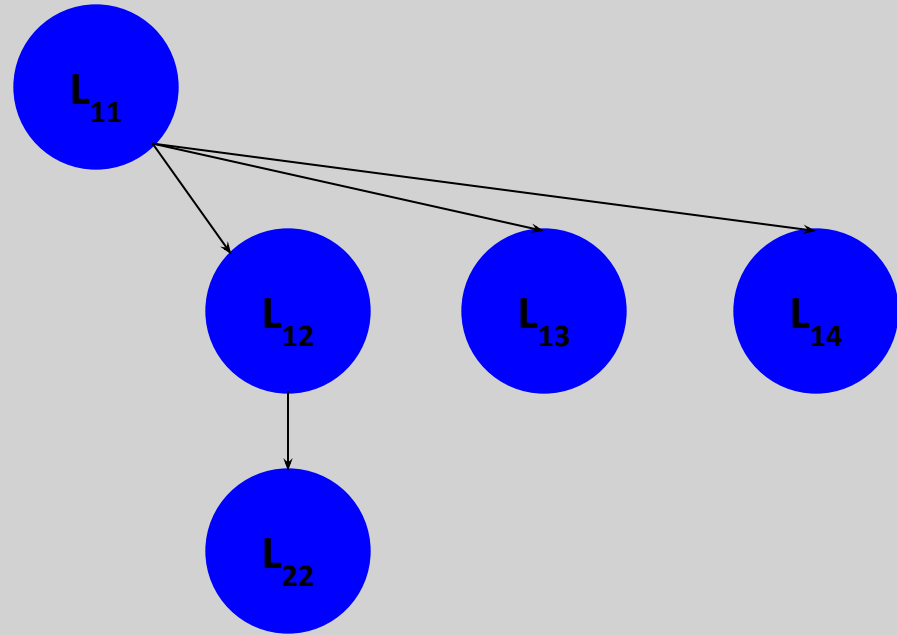
L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)



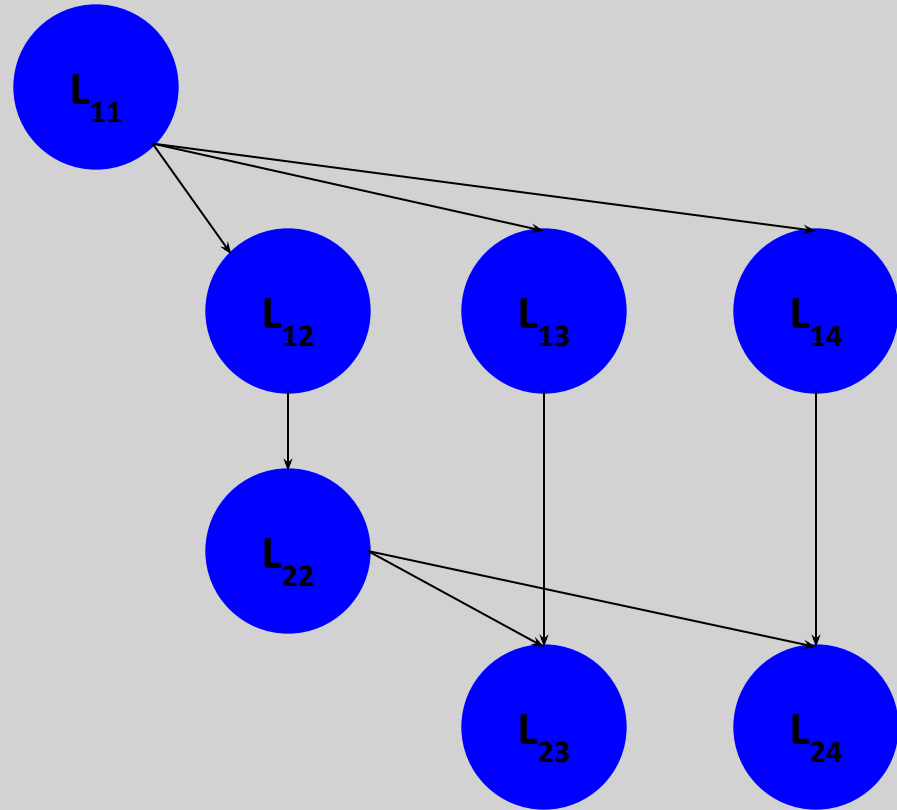
L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)



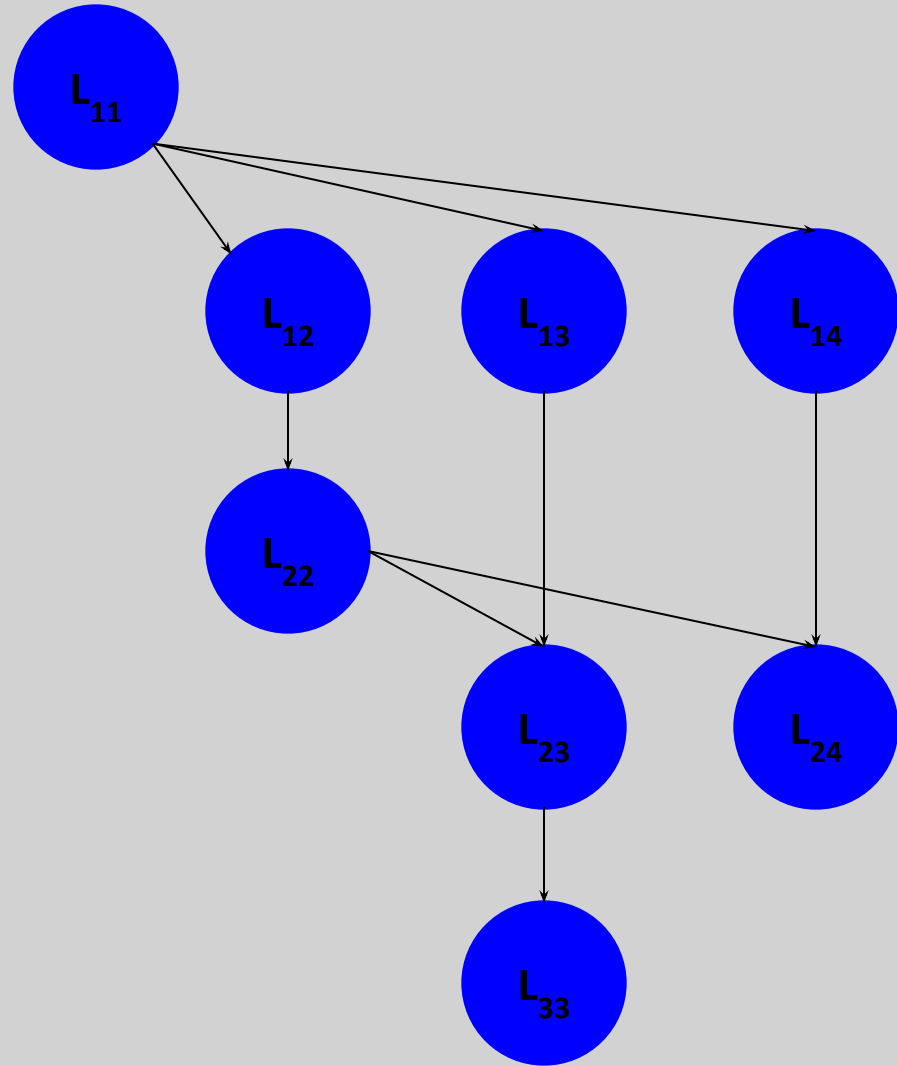
L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)



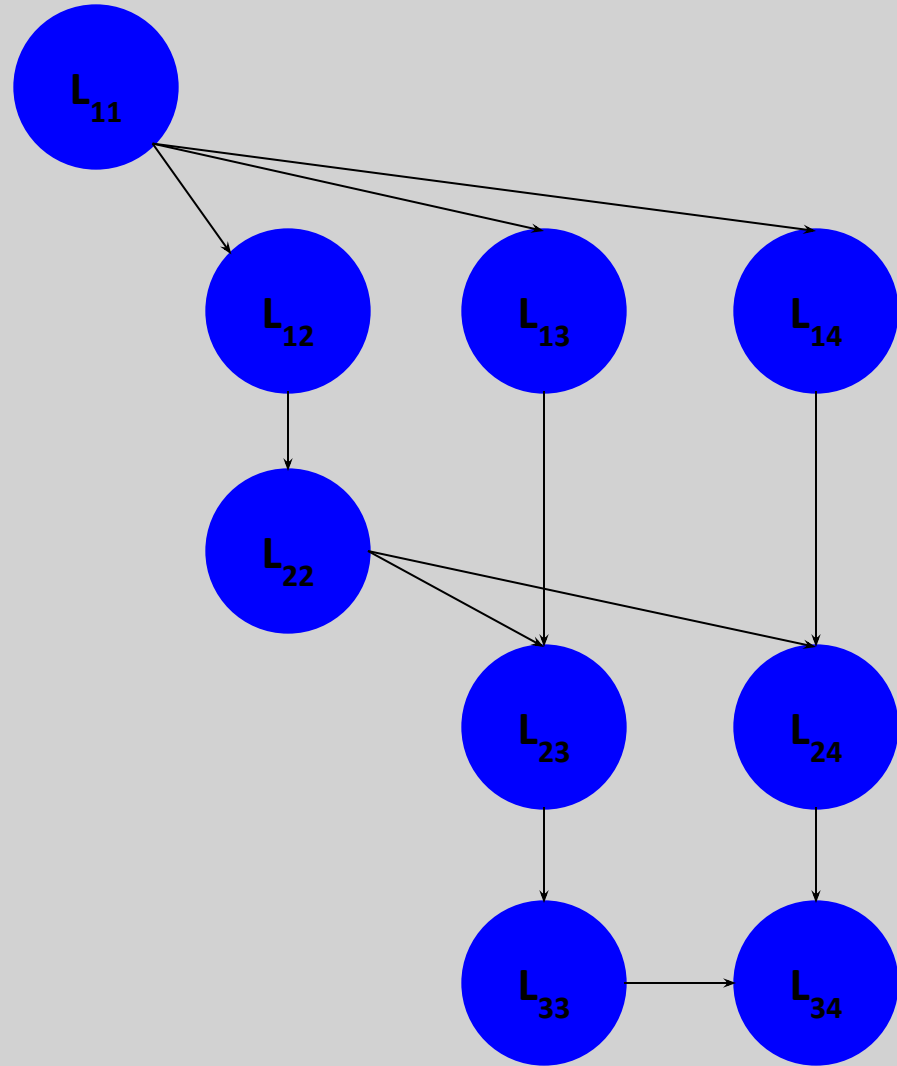
L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)



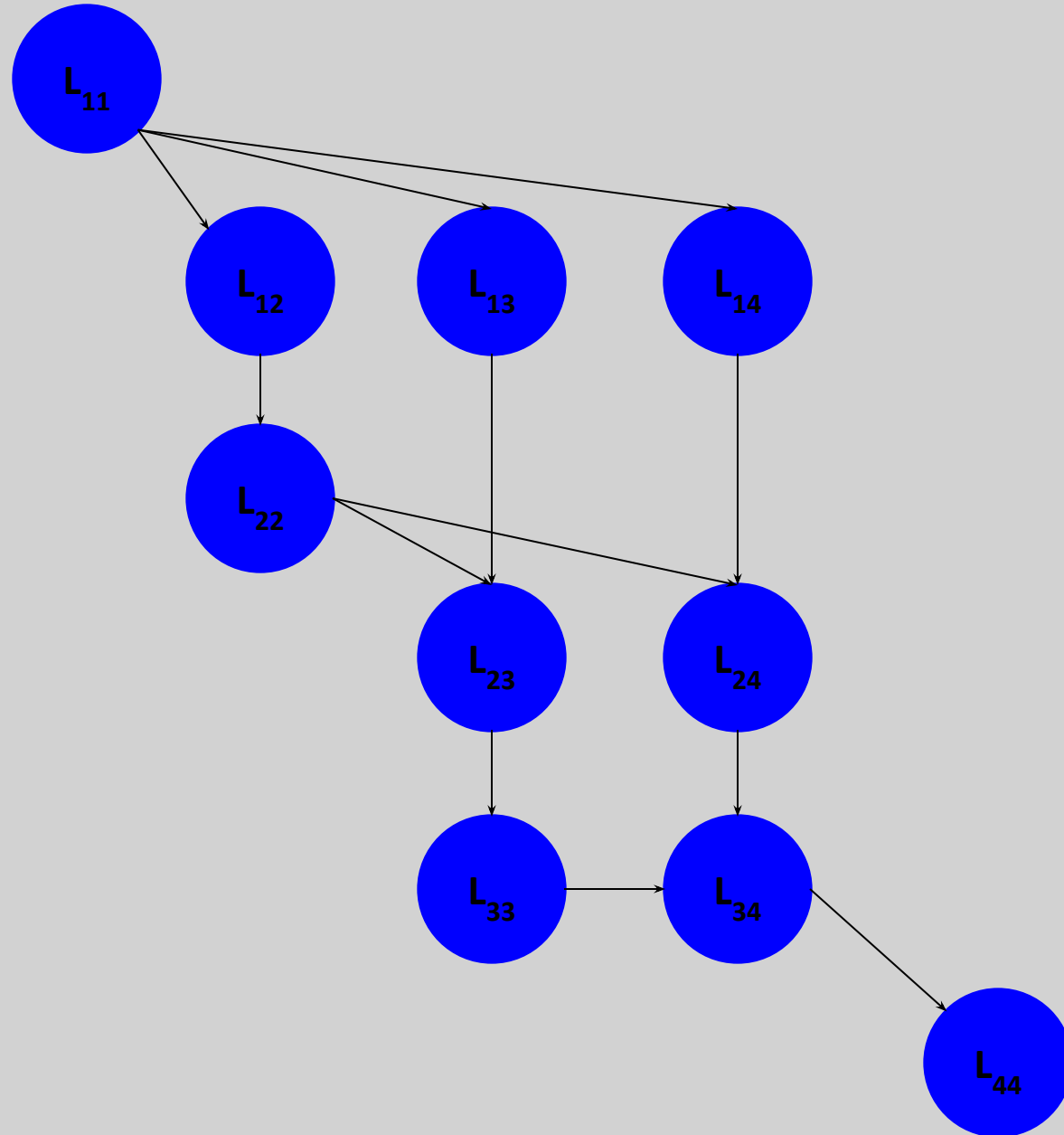
L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)

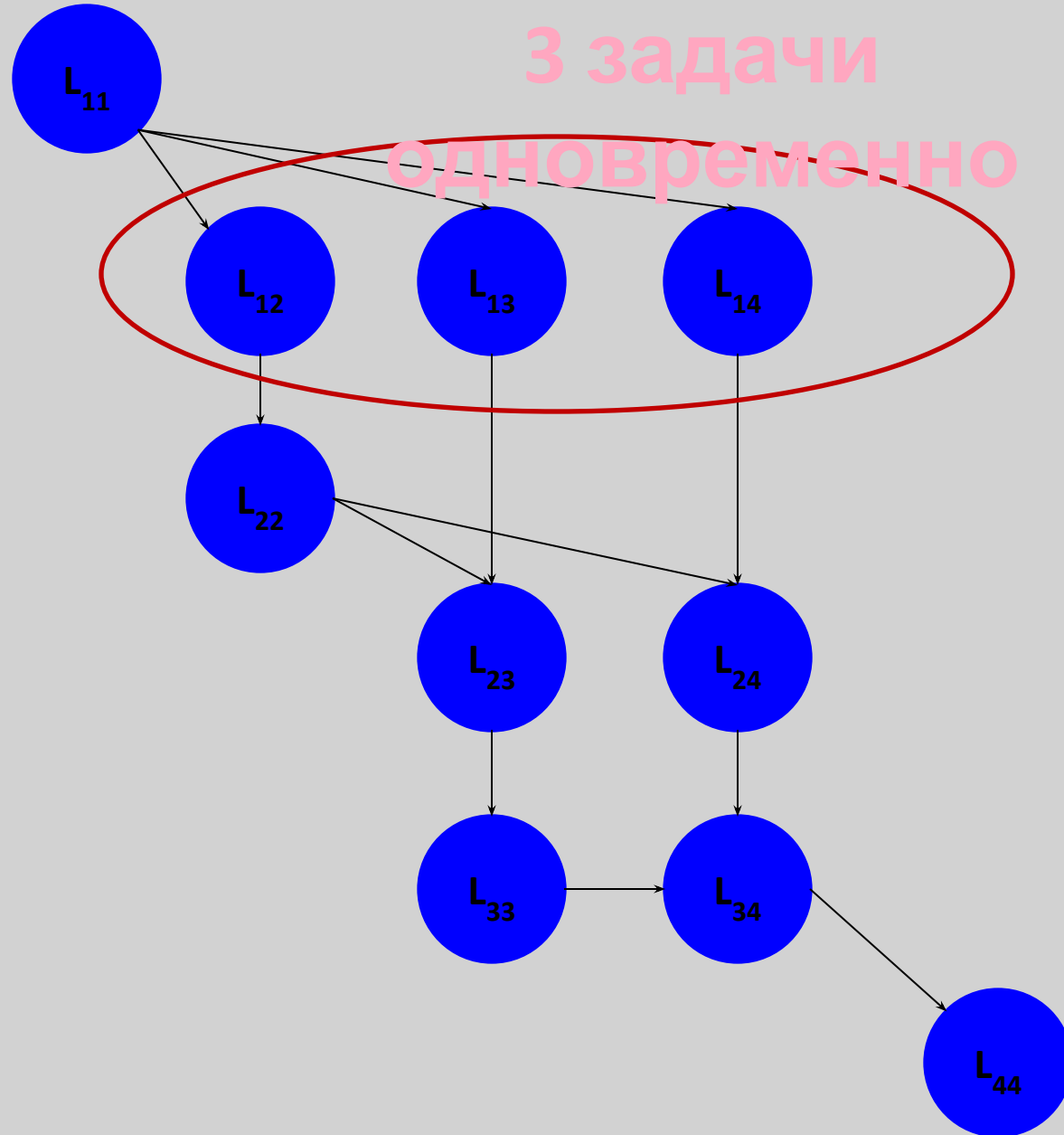


L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Дag подход (Directed acyclic graph)

3 задачи

одновременно



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

Более того, мы можем постепенно модифицировать L_{ij} , а не только перед самым вычислением L_{ij} , т.е. добавляется дополнительная возможность разбить работу

$$\tilde{L}_{34} = L_{34} - L_{14} \cdot L_{13}$$

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

- Перемножение прямоугольных матриц *GEMM
- Обращение треугольной матрицы на прямоугольной *TRSM
- Разложение квадратной матрицы на произведение треугольных

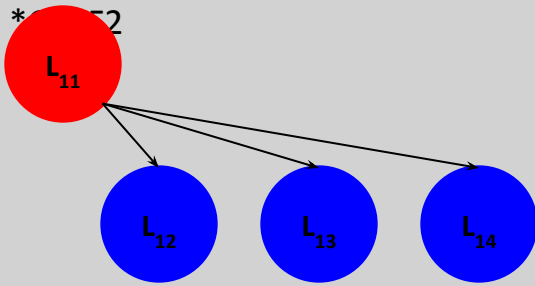
*GEMM

L_{11}

L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}




Dag подход (Directed acyclic graph)

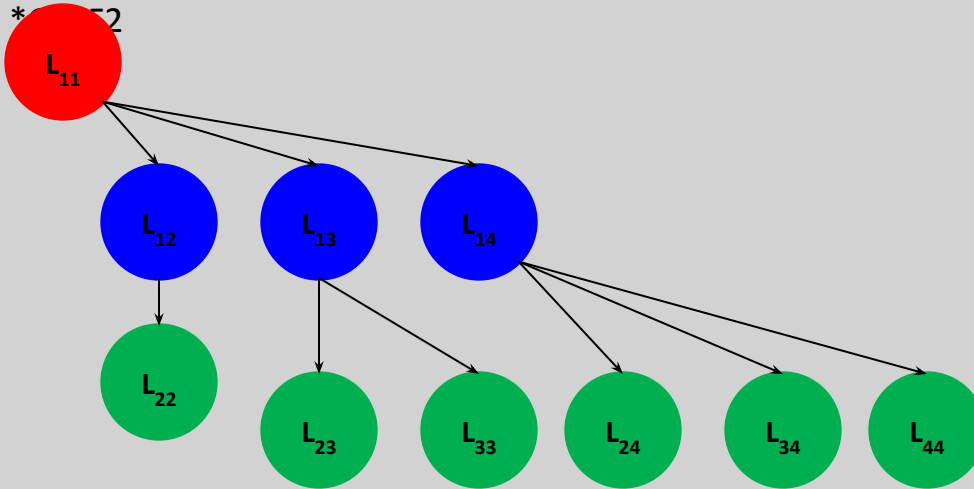
- Перемножение прямоугольных матриц *GEMM
- Обращение треугольной матрицы на прямоугольной *TRSM
- Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}




Dag подход (Directed acyclic graph)

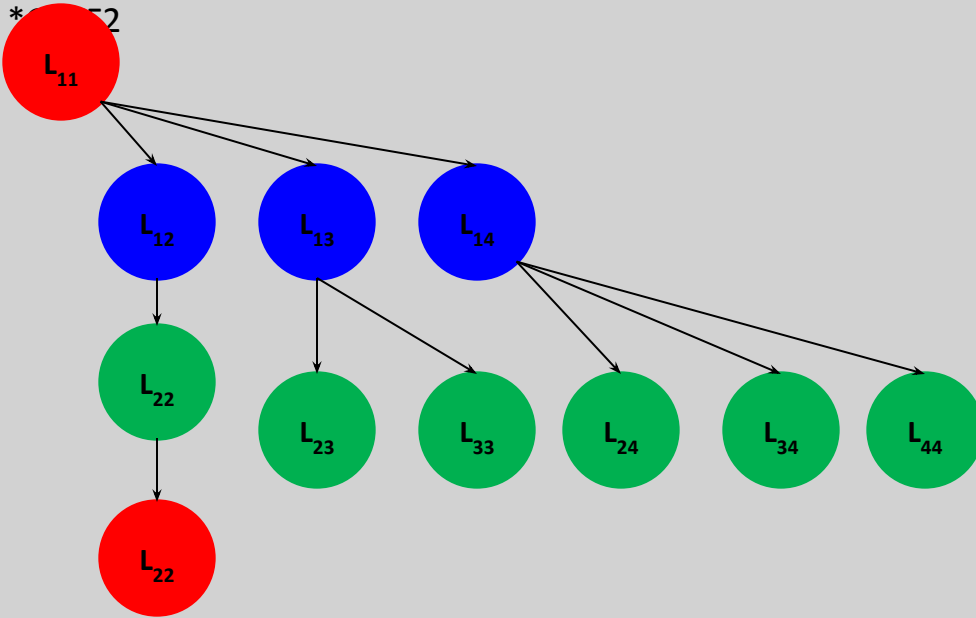
-  Перемножение прямоугольных матриц *GEMM
-  Обращение треугольной матрицы на прямоугольной *TRSM
-  Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}




Dag подход (Directed acyclic graph)

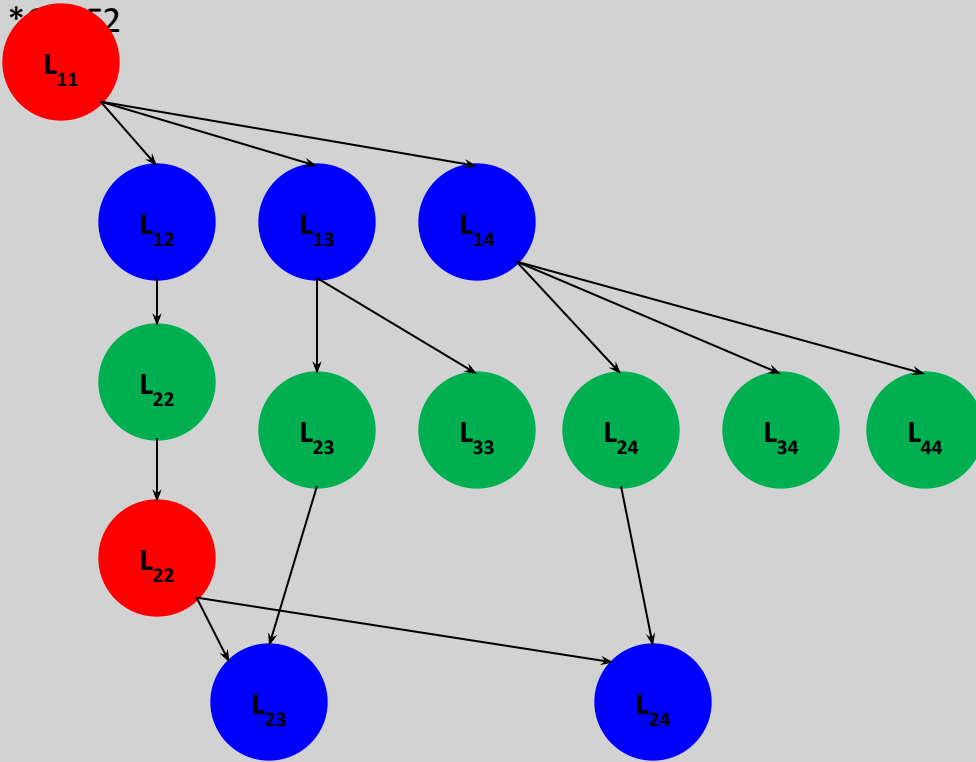
-  Перемножение прямоугольных матриц *GEMM
-  Обращение треугольной матрицы на прямоугольной *TRSM
-  Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}




Dag подход (Directed acyclic graph)

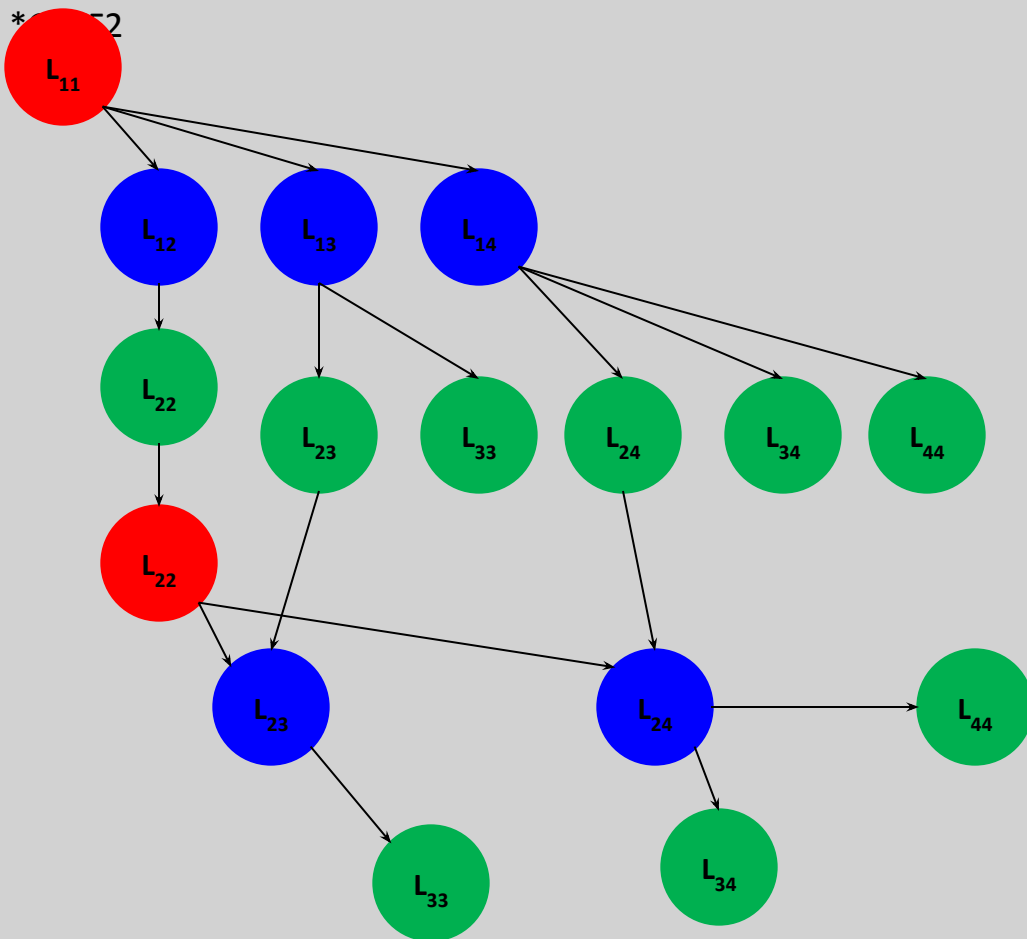
-  Перемножение прямоугольных матриц *GEMM
-  Обращение треугольной матрицы на прямоугольной *TRSM
-  Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}




Dag подход (Directed acyclic graph)

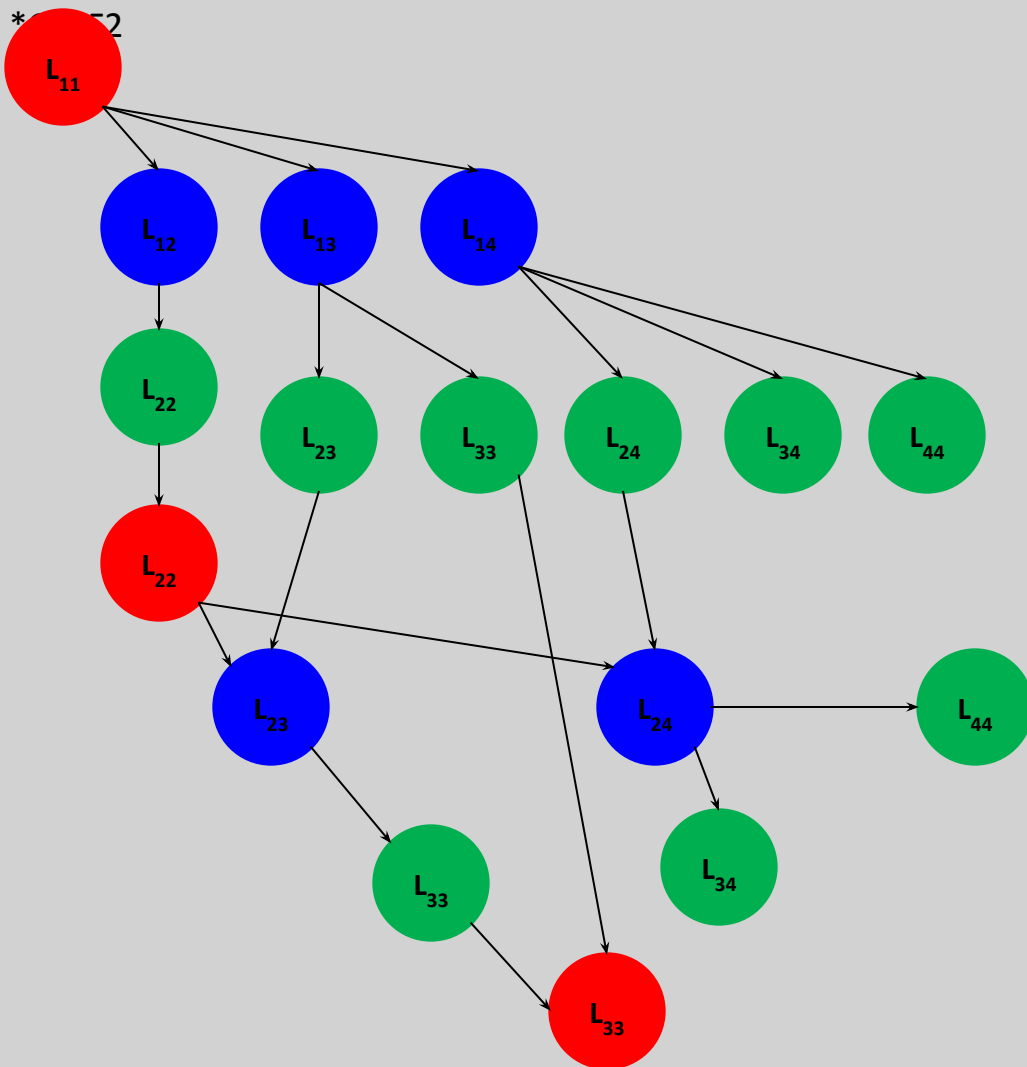
-  Перемножение прямоугольных матриц *GEMM
-  Обращение треугольной матрицы на прямоугольной *TRSM
-  Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

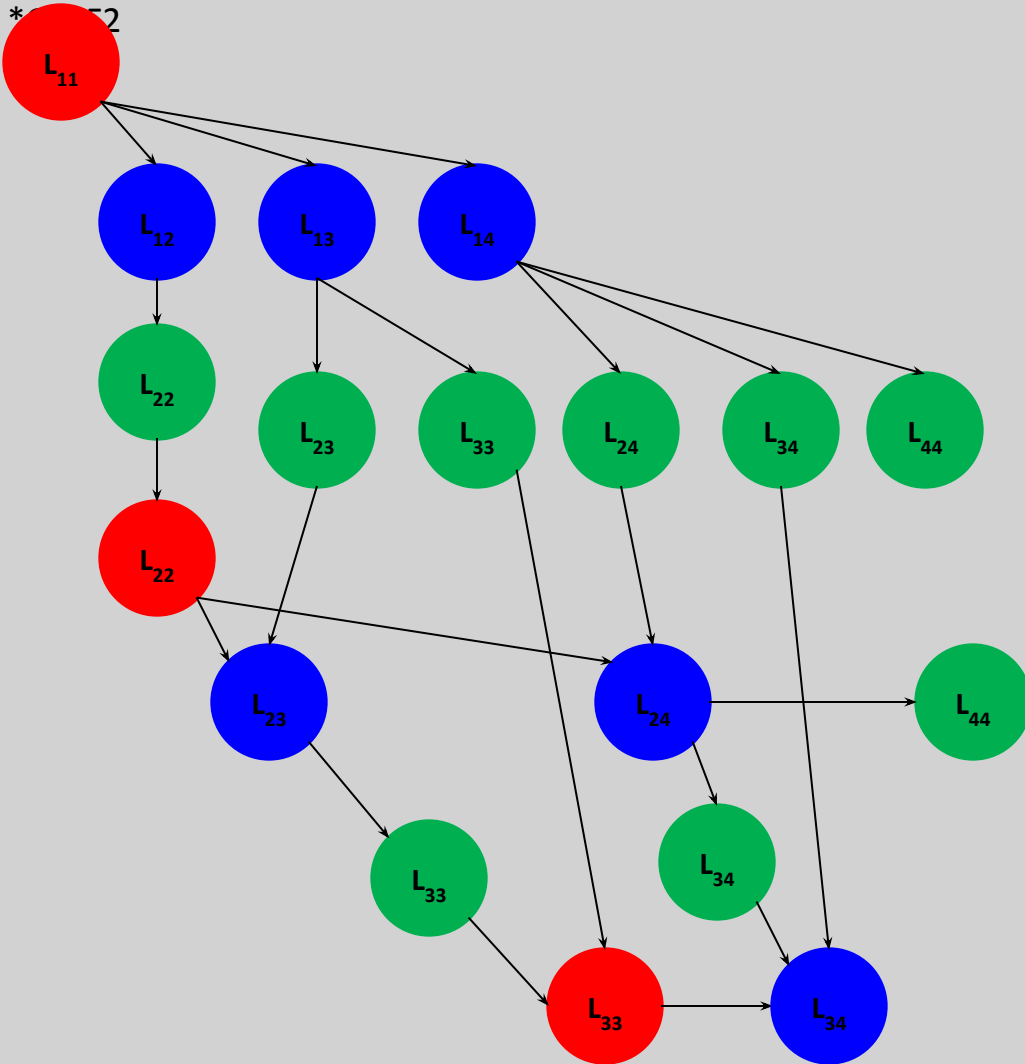
-  Перемножение прямоугольных матриц *GEMM
-  Обращение треугольной матрицы на прямоугольной *TRSM
-  Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

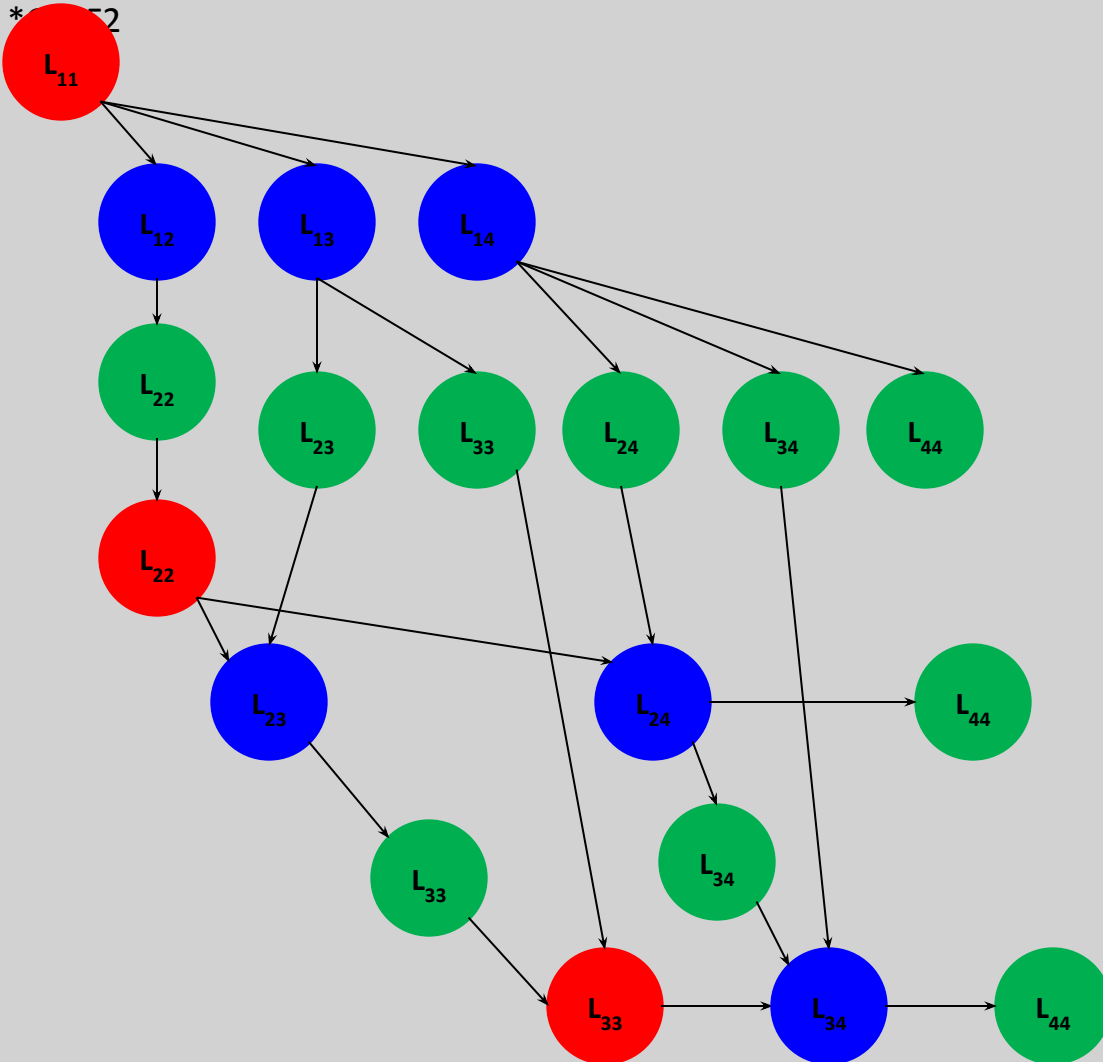
- Перемножение прямоугольных матриц *GEMM
- Обращение треугольной матрицы на прямоугольной *TRSM
- Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

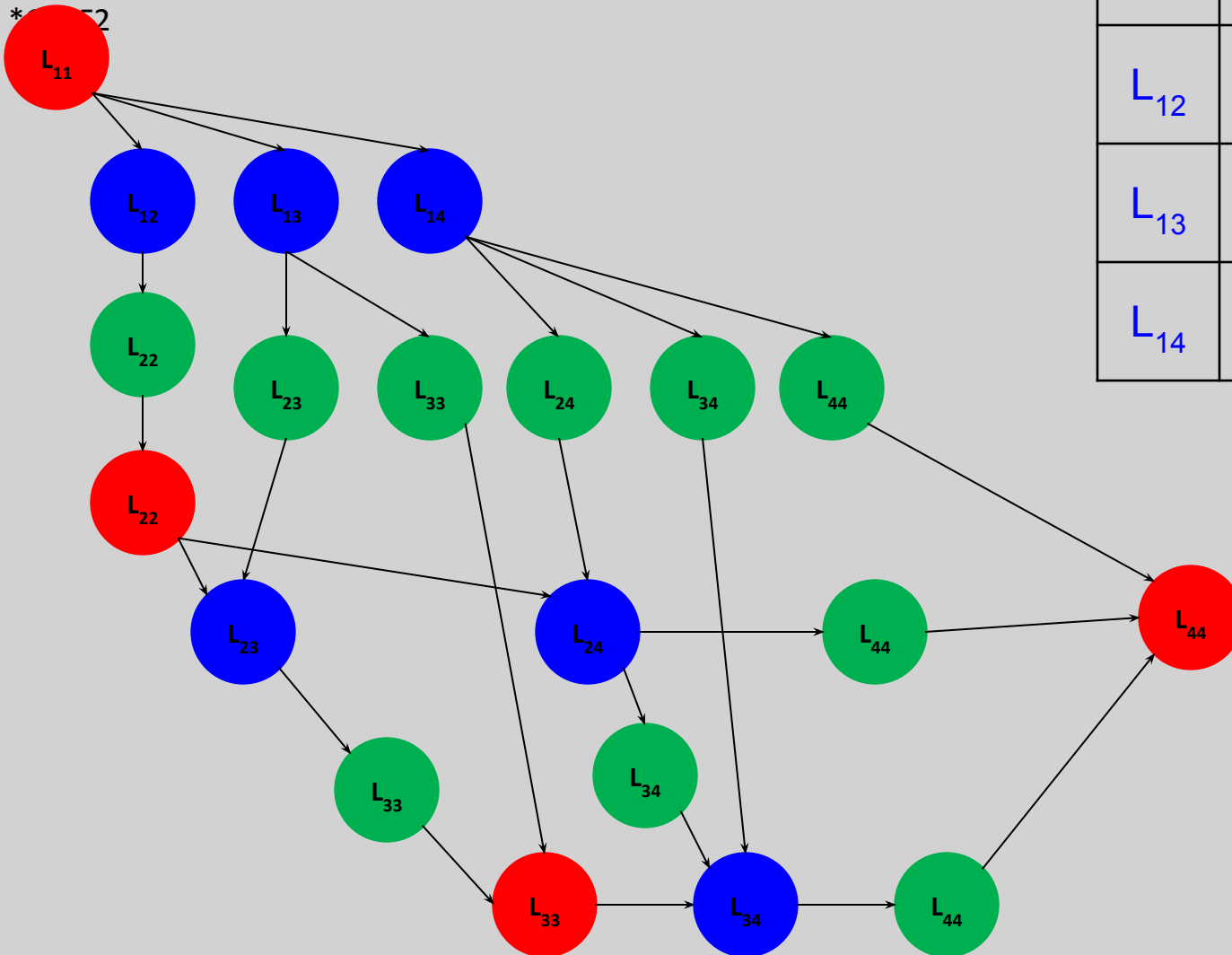
- Перемножение прямоугольных матриц *GEMM
- Обращение треугольной матрицы на прямоугольной *TRSM
- Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

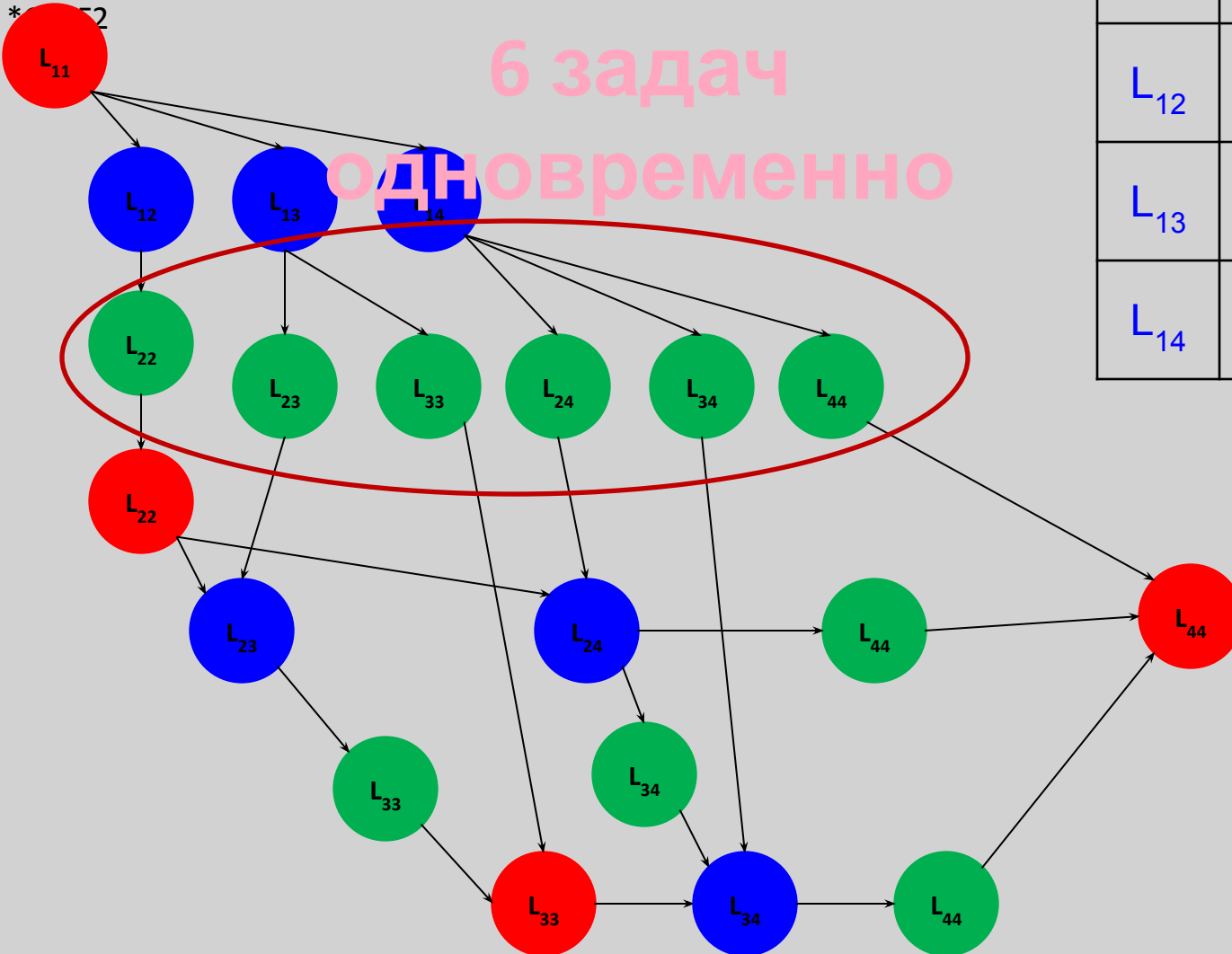
- Перемножение прямоугольных матриц *GEMM
- Обращение треугольной матрицы на прямоугольной *TRSM
- Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

- Перемножение прямоугольных матриц *GEMM
- Обращение треугольной матрицы на прямоугольной *TRSM
- Разложение квадратной матрицы на произведение треугольных



L_{11}	0	0	0
L_{12}	L_{22}	0	0
L_{13}	L_{23}	L_{33}	0
L_{14}	L_{24}	L_{34}	L_{44}

Dag подход (Directed acyclic graph)

Плюсы:

- Очень хорошая масштабируемость на старте алгоритма
- Динамическое распределение задач
- Возможность изменения размеров блоков в зависимости от положения в графе

Минусы:

- Слабая масштабируемость на окончании алгоритма
- Динамическое распределение задач

Далее...

- Как реализовать алгоритм для очень большого числа ядер (> 100)?
- Как модифицировать алгоритм в случае большого количества кэшей разного уровня?
- Как выбирать размер блоков в зависимости от процессора/платформы?

Вопросы открыты.....

Вопросы и Ответы

