

## Лекция 3.

# Оценка коммуникационной трудоемкости параллельных алгоритмов



# Содержание

---

- ❑ Общая характеристика механизмов передачи данных
  - Алгоритмы маршрутизации
  - Методы передачи данных
- ❑ Анализ трудоемкости основных операций передачи данных
- ❑ Методы логического представления топологии коммуникационной среды
- ❑ Оценка трудоемкости операций передачи данных для кластерных систем



# Введение

---

- Данный раздел посвящен вопросам анализа информационных потоков, возникающих при выполнении параллельных алгоритмов:
  - дается общая характеристика механизмов передачи данных,
  - проводится анализ трудоемкости основных операций обмена информацией,
  - рассматриваются методы логического представления структуры МВС.

*Затраты на организацию взаимодействия при помощи передачи сообщений существенным образом влияют на эффективность параллельных вычислений*



# Общая характеристика механизмов передачи данных...

---

- **Алгоритмы маршрутизации** определяют путь передачи данных от процессора-источника сообщения до процессора, к которому сообщение должно быть доставлено:
  - *оптимальные*, определяющие всегда наикратчайшие пути передачи данных, и *неоптимальные* алгоритмы маршрутизации,
  - *детерминированные* и *адаптивные* методы выбора маршрутов (адаптивные алгоритмы определяют пути передачи данных в зависимости от существующей загрузки коммуникационных каналов).



# Общая характеристика механизмов передачи данных...

## □ Алгоритмы маршрутизации

– *метод по координатной маршрутизации (dimension-ordered routing)* – один из самых распространенных оптимальных методов маршрутизации:

- Поиск путей передачи данных осуществляется поочередно для каждой размерности топологии сети коммуникации,
- Для двумерной решетки: передача данных сначала выполняется по одному направлению, а затем данные передаются вдоль другого направления (*алгоритм XY-маршрутизации*),
- Для гиперкуба: циклическая передача данных процессору, определяемому первой различающейся битовой позицией в номерах процессоров, на котором сообщение располагается в данный момент времени и на который сообщение должно быть передано.



# Общая характеристика механизмов передачи данных...

## □ Методы передачи данных...

Время передачи данных между процессорами определяет коммуникационную составляющую (*communication latency*) длительности выполнения параллельного алгоритма.

Основной набор параметров, используемый при оценке времени передачи данных, включает:

- **время начальной подготовки** ( $t_n$ ) характеризует длительность подготовки сообщения для передачи, поиска маршрута в сети и т.п.,
- **время передачи служебных данных** ( $t_c$ ) между двумя соседними процессорами (т.е. для процессоров, между которыми имеется физический канал передачи данных); к служебным данным может относиться заголовок сообщения, блок данных для обнаружения ошибок передачи и т.п.,
- **время передачи одного слова данных** по одному каналу передачи данных ( $t_k$ ); длительность подобной передачи определяется полосой пропускания коммуникационных каналов в сети.



# Общая характеристика механизмов передачи данных...

---

## □ Методы передачи данных...

*Метод передачи сообщений (МПС)* осуществляет передачу данных как неделимых (атомарных) блоков информации (*store-and-forward routing or SFR*):

- процессор, содержащий сообщение для передачи, готовит весь объем данных для передачи, определяет процессор, которому следует направить данные, и запускает операцию пересылки данных,
- процессор, которому направлено сообщение, в первую очередь осуществляет прием полностью всех пересылаемых данных и только затем приступает к пересылке принятого сообщения далее по маршруту.



# Общая характеристика механизмов передачи данных...

---

## □ Методы передачи данных...

*Метод передачи пакетов (МПП)* основан на представлении пересылаемых сообщений в виде блоков информации меньшего размера (*cut-through routing* or *CTR*):

- принимающий процессор может осуществлять пересылку данных по дальнейшему маршруту непосредственно сразу после приема очередного пакета, не дожидаясь завершения приема данных всего сообщения.





# Общая характеристика механизмов передачи данных...

---

## □ Методы передачи данных...

Время пересылки данных  $t_{пд}$  размером  $m$  байт по маршруту длиной  $l$  определяется выражением:

– Для метода передачи сообщений:

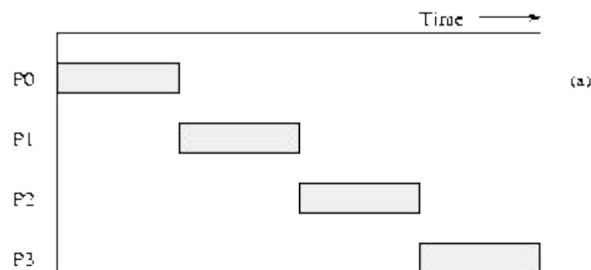
при достаточно длинных сообщениях временем пересылки служебных данных можно пренебречь:

– Для метода передачи пакетов:

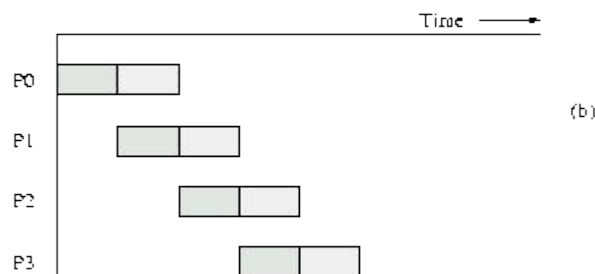


# Общая характеристика механизмов передачи данных...

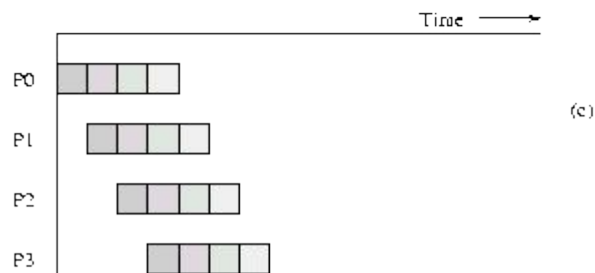
## □ Методы передачи данных...



□ Метод передачи сообщений



□ Метод пересылки пакетов  
(сообщение разбивается на 2 пакета)



□ Метод пересылки пакетов  
(сообщение разбивается на 4 пакета)

# Общая характеристика механизмов передачи данных

---

- Метод передачи пакетов (оценка применимости):
  - приводит к более быстрой пересылке данных,
  - снижает потребность в памяти для хранения пересылаемых данных для организации приема-передачи сообщений,
  - для передачи могут использоваться одновременно разные коммуникационные каналы,
  - требует разработки более сложного аппаратного и программного обеспечения сети,
  - может увечить накладные расходы (время подготовки и время передачи служебных данных),
  - при передаче пакетов возможно возникновение конфликтных ситуаций (*дедлоков*).



# Анализ трудоемкости основных операций передачи данных...

---

- При анализе параллельных способов решения сложных научно-технических задач могут быть выделены основные коммуникационные действия, которые или наиболее широко распространены в практике, или к которым могут быть сведены многие другие процессы приема-передачи сообщений,
- Для большинства операций коммуникации существуют процедуры, обратные по действию исходным операциям (так, например, операции передачи данных от одного процессора всем имеющимся процессорам сети соответствует операция приема в одном процессоре сообщений от всех остальных процессоров).



# Анализ трудоемкости основных операций передачи данных...

## □ Передача данных между двумя процессорами сети

Трудоемкость данной коммуникационной операции может быть получена путем подстановки длины максимального пути в выражения для времени передачи данных при разных методах коммуникации.

Топология	Передача сообщений	Передача пакетов
Кольцо		
Решетка-тор		
Гиперкуб		



# Анализ трудоемкости основных операций передачи данных...

---

- **Передача данных от одного процессора всем остальным процессорам сети...**

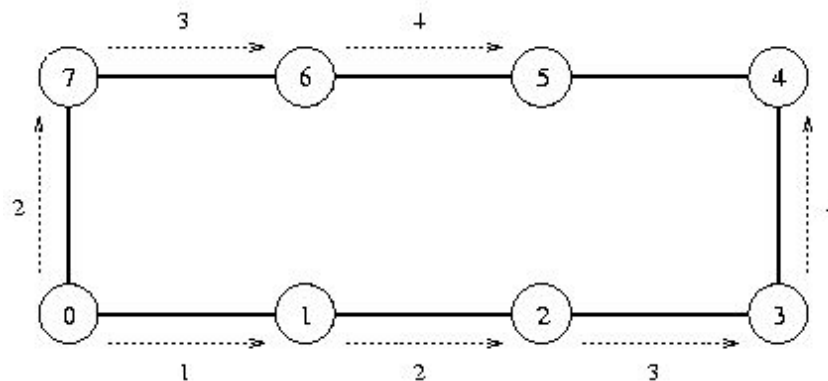
Операция передачи данных (одного и того же сообщения) от одного процессора всем остальным процессорам сети (*one-to-all broadcast or single-node broadcast*) является одним из наиболее часто выполняемых коммуникационных действий; двойственная операция передачи – прием на одном процессоре сообщений от всех остальных процессоров сети (*single-node accumulation*).



# Анализ трудоемкости основных операций передачи данных...

- **Передача данных от одного процессора всем остальным процессорам сети (*передача сообщений*)...**

Для **кольцевой топологии** процессор-источник рассылки может инициировать передачу данных сразу двум своим соседям, которые, в свою очередь, приняв сообщение, организуют пересылку далее по кольцу:

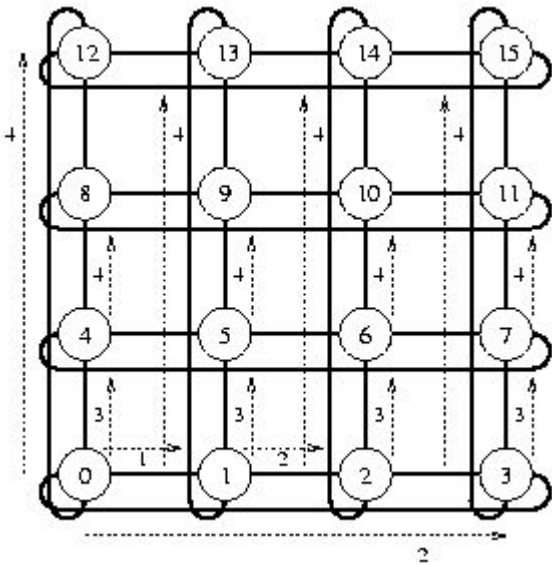


Трудоемкость выполнения операции рассылки в этом случае будет определяться соотношением:

# Анализ трудоемкости основных операций передачи данных...

- ❑ Передача данных от одного процессора всем остальным процессорам сети (*передача сообщений*)...

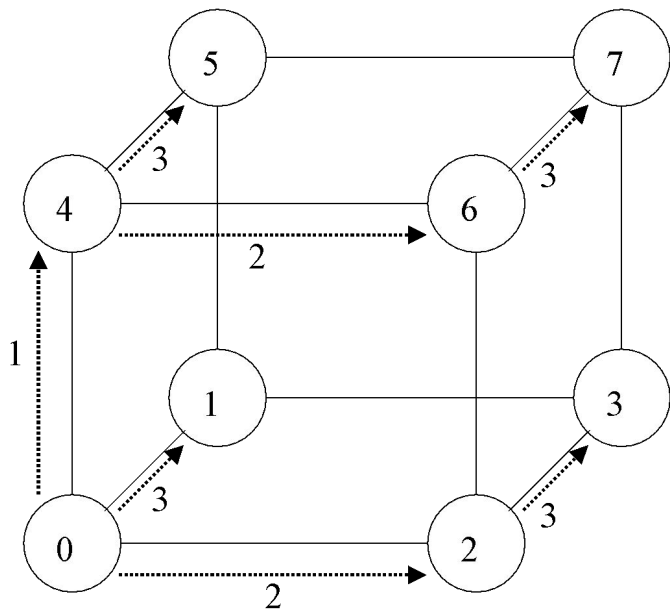
Для топологии типа **решетки-тора** рассылка может быть выполнена в виде двухэтапной процедуры. На первом этапе организуется передача сообщения всем процессорам сети, располагающимся на той же горизонтали решетки, что и процессор-инициатор передачи; на втором этапе процессоры, получившие копию данных на первом этапе, рассылают сообщения по своим соответствующим вертикалям. Длительности операции рассылки в соответствии с описанным алгоритмом определяется соотношением:





# Анализ трудоемкости основных операций передачи данных...

- ❑ Передача данных от одного процессора всем остальным процессорам сети (*передача сообщений*)



Для гиперкуба рассылка может быть выполнена в ходе  $N$ -этапной процедуры передачи данных. На первом этапе процессор-источник сообщения передает данные одному из своих соседей – в результате после первого этапа имеется два процессора, имеющих копию пересылаемых данных. На втором этапе два процессора, задействованные на первом этапе, пересылают сообщение своим соседям по второй размерности и т.д.

В результате такой рассылки время операции оценивается при помощи выражения

# Анализ трудоемкости основных операций передачи данных...

---

- **Передача данных от одного процессора всем остальным процессорам сети (*передача пакетов*)...**

Для топологии типа **кольца** алгоритм рассылки может быть получен путем логического представления кольцевой структуры сети в виде гиперкуба. В результате на этапе рассылки процессор-источник сообщения передает данные процессору, находящемуся на расстоянии  $p/2$  от исходного процессора. Далее, на втором этапе оба процессора, уже имеющие рассылаемые данные после первого этапа, передают сообщения процессорам, находящиеся на расстоянии  $p/4$  и т.д.

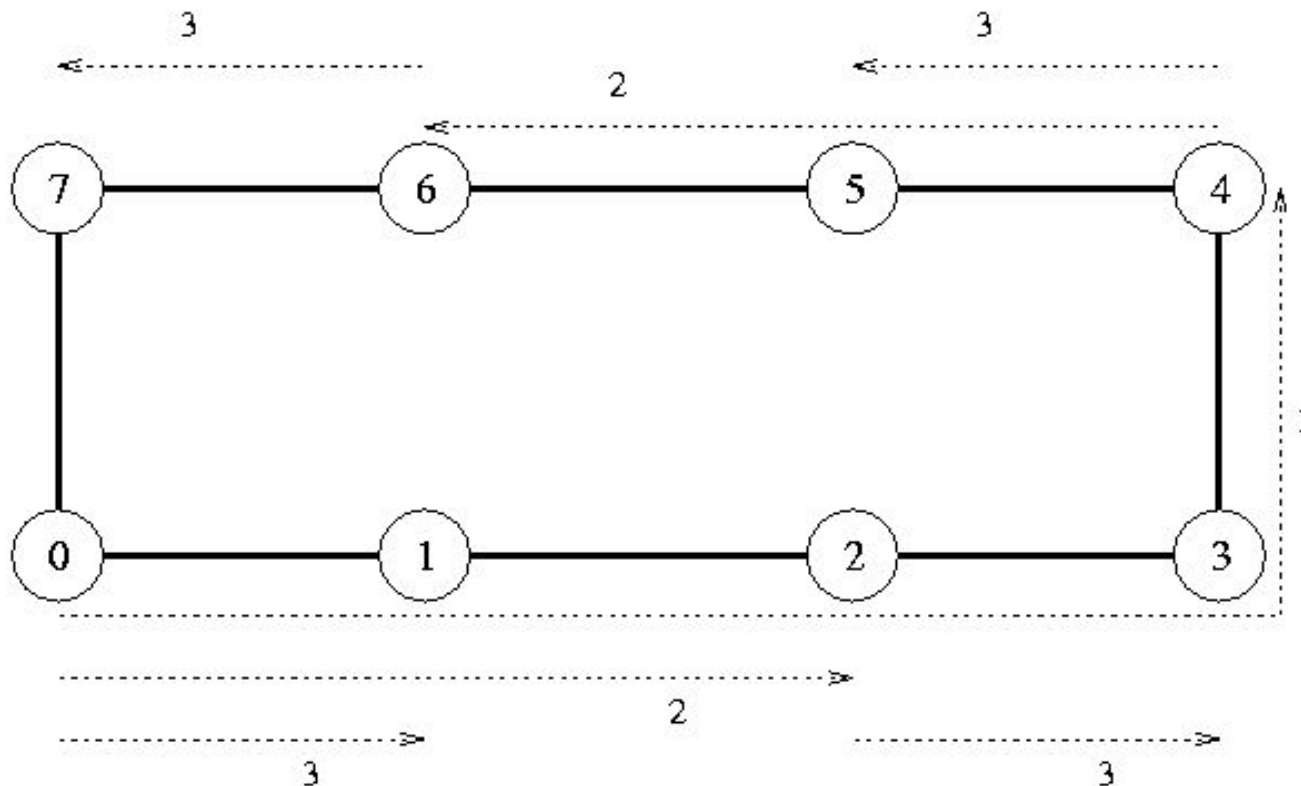
Трудоемкость выполнения операции рассылки при таком методе передачи данных определяется соотношением:



# Анализ трудоемкости основных операций передачи данных...

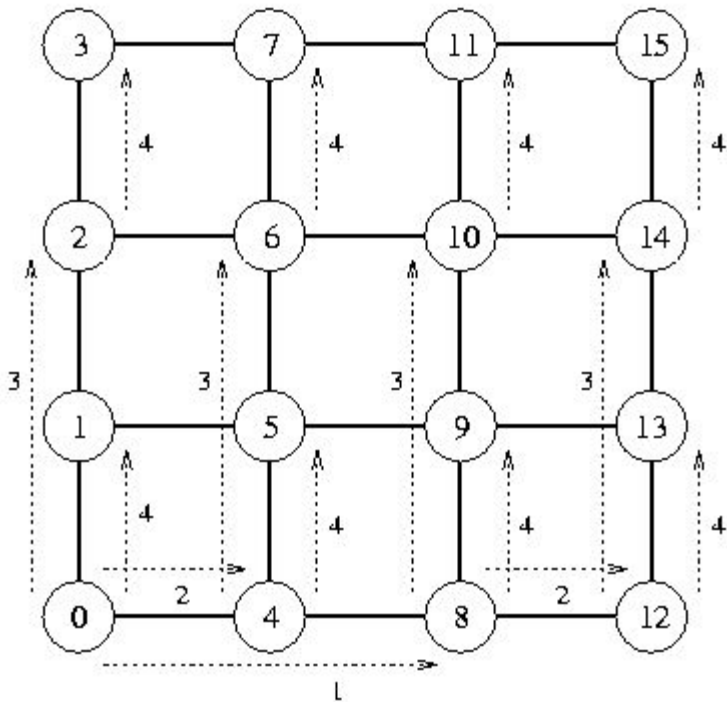
- ❑ Передача данных от одного процессора всем остальным процессорам сети (*передача пакетов*)...

Топология типа кольца



# Анализ трудоемкости основных операций передачи данных...

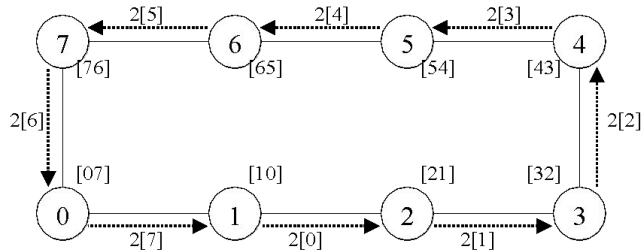
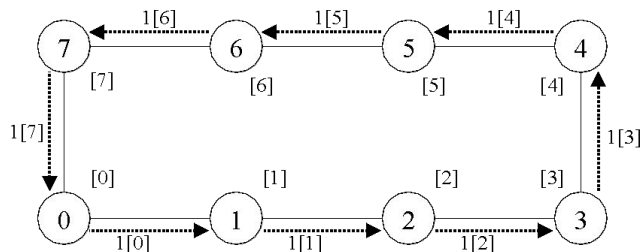
- Передача данных от одного процессора всем остальным процессорам сети (*передача пакетов*)



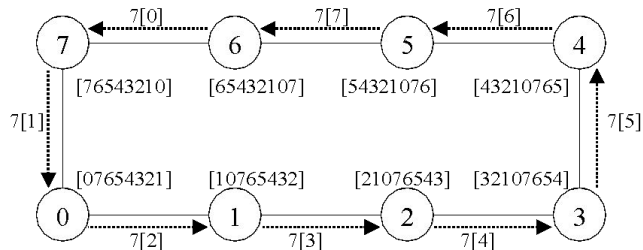
Для топологии типа **решетки-тора** алгоритм рассылки может быть получен из способа передачи данных, примененного для кольцевой структуры сети, в соответствии с тем же способом обобщения, что и в случае использования метода передачи сообщений. Получаемый в результате такого обобщения алгоритм рассылки характеризуется следующим соотношением для оценки времени выполнения:

# Анализ трудоемкости основных операций передачи данных...

- Передача данных от всех процессоров всем процессорам сети (*передача сообщений*)...



...



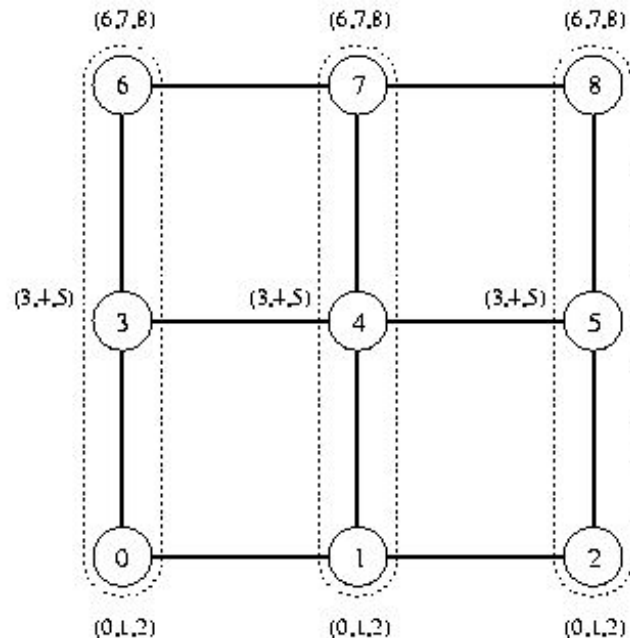
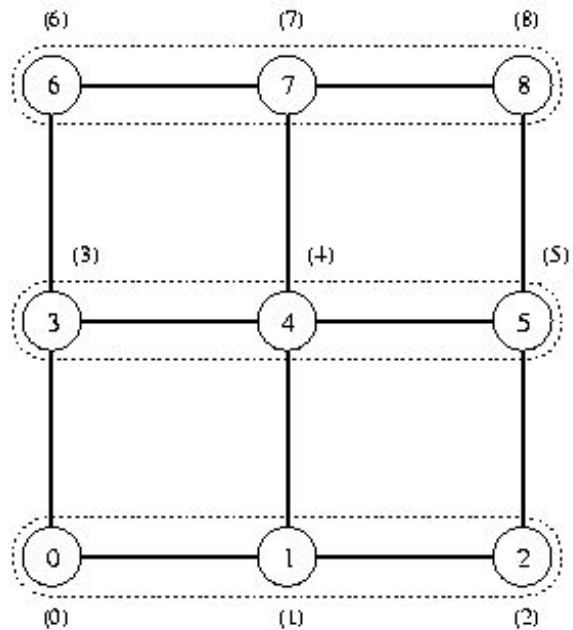
Для кольцевой топологии каждый процессор может инициировать рассылку своего сообщения одновременно (в каком-либо выбранном направлении по кольцу). В любой момент времени каждый процессор выполняет прием и передачу данных; завершение операции множественной рассылки произойдет через  $(p-1)$  цикл передачи данных.

Длительность выполнения операции рассылки оценивается соотношением:



# Анализ трудоемкости основных операций передачи данных...

- Передача данных от всех процессоров всем процессорам сети (*передача сообщений*)...



**Решетка-тор** - общая длительность операции рассылки определяется соотношением



# Анализ трудоемкости основных операций передачи данных...

---

- Передача данных от всех процессоров всем процессорам сети (*операция редукция*)...

Широко распространенный пример операции множественной рассылки - *задача редукции (reduction)* или, другими словами, процедура выполнения той или иной обработки данных, получаемых на каждом процессоре в ходе множественной рассылки (например, проблема вычисления суммы значений, находящихся на разных процессорах, и рассылки полученной суммы по всем процессорам сети).



# Анализ трудоемкости основных операций передачи данных...

## □ Передача данных от всех процессоров всем процессорам сети (*операция редукции*)

Способы решения задачи редукции могут состоять в следующем:

- **непосредственный подход**: выполнение операции множественной рассылки и последующая обработка данных на каждом процессоре,
- **более эффективный алгоритм**: операция одиночного приема данных на отдельном процессоре, выполнение на этом процессоре действий по обработке данных, и рассылка полученного результата обработки всем процессорам сети,
- **наилучший способ** - совмещение процедуры множественной рассылки и действий по обработке данных, когда каждый процессор сразу же после приема очередного сообщения реализует требуемую обработку полученных данных. При этом время решения задачи ( при топологии сети в виде гиперкуба и размере сообщения  $m=1$ ):

$$t_{nd} = (t_n + t_k) \log_2 p$$





# Анализ трудоемкости основных операций передачи данных...

---

## □ Передача данных от всех процессоров всем процессорам сети

Другим типовым примером использования операции множественной рассылки является **задача нахождения частных сумм** последовательности значений (в литературе эта задача известна под названием *prefix sum problem*):

Алгоритм решения данной задачи также может быть получен при помощи конкретизации общего способа выполнения множественной операции рассылки, когда процессор выполняет суммирование полученного значения (но только в том случае, если процессор-отправитель значения имеет меньший номер, чем процессор-получатель).

---



# Методы логического представления топологии коммуникационной среды...

---

- Способы логического представления (отображения) топологий характеризуются следующими тремя основными характеристиками:
  - **уплотнение дуг** (*congestion*), выражаемое как максимальное количество дуг логической топологии, отображаемых в одну линию передачи физической топологии,
  - **удлинение дуг** (*dilation*), определяемое как путь максимальной длины физической топологии, на который отображаемая дуга логической топологии,
  - **увеличение вершин** (*expansion*), вычисляемое как отношение количества вершин в логической и физической топологиях.



# Методы логического представления топологии коммуникационной среды...

## □ Представление кольцевой топологии в виде гиперкуба...

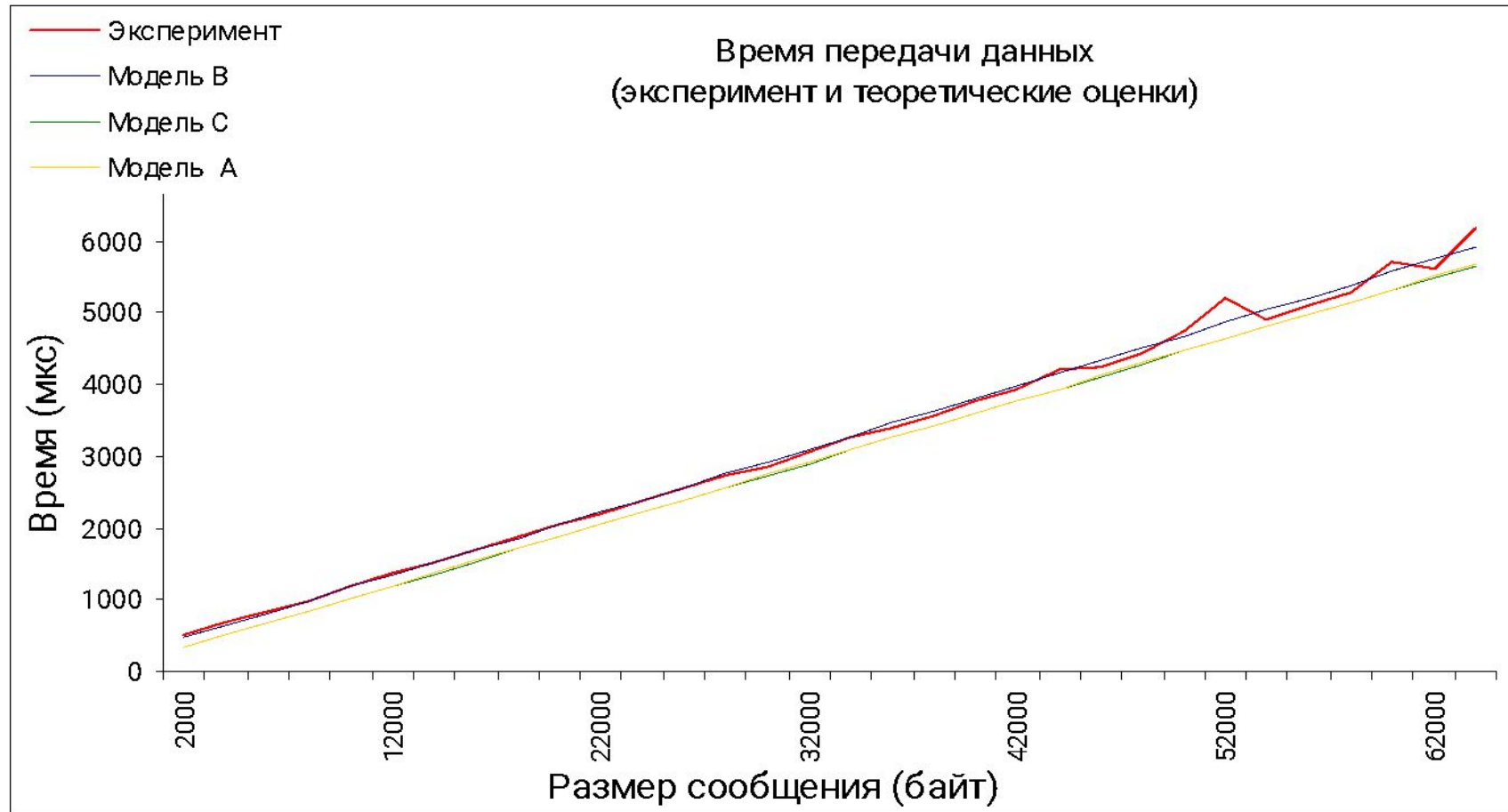
Отображение кольцевой топологии на гиперкуб для сети из  $p=8$  процессоров:

Код Грея для N=1	Код Грея для N=2	Код Грея для N=3	Номера процессоров	
			гиперкуба	кольца
0	00	000	0	0
1	01	001	1	1
	11	011	3	2
	10	010	2	3
		110	6	4
		111	7	5
		101	5	6
		100	4	7



# Оценка трудоемкости операций передачи данных для кластерных систем...

## □ Результаты вычислительных экспериментов...



# Оценка трудоемкости операций передачи данных для кластерных систем...

## □ Результаты вычислительных экспериментов

Объем сообщения (байт)	Время передачи (мкс)	Погрешность теоретической оценки времени передачи данных, в %		
		Модель А	Модель В	Модель С
2000	495	33.45%	7.93%	34.80%
10000	1184	13.91%	1.70%	14.48%
20000	2055	8.44%	0.44%	8.77%
30000	2874	4.53%	-1.87%	4.76%
40000	3758	4.04%	-1.38%	4.22%
50000	4749	5.91%	1.21%	6.05%
60000	5730	6.97%	2.73%	7.09%



# Заключение...

---

- ❑ Представлена общая характеристика алгоритмов маршрутизации и методов передачи данных. Для подробного рассмотрения выделены метод передачи сообщений (*store-and-forward routing*) и метод передачи пакетов (*cut-through routing*), для которых определены оценки времени выполнения коммуникационных операций.
- ❑ Определены основные типы операций передачи данных, выполняемых в ходе параллельных вычислений. Для всех операций рассмотрены алгоритмы их выполнения на примере топологий кольца, решетки и гиперкуба. Приведены оценки их временной трудоемкости как для метода передачи сообщений, так и для метода передачи пакетов.



# Заключение

---

- ❑ Рассмотрены *методы логического представления топологий* на основе конкретных (физических) межпроцессорных структур.
- ❑ Проведен анализ моделей, при помощи которых могут быть получены оценки времени выполнения операций передачи данных для кластерных вычислительных систем. По результатам сравнения для дальнейшего использования при оценке временной трудоемкости коммуникационных операций выбрана наиболее простая модель - *модель Хокни*.



# Вопросы для обсуждения

---

- ❑ Оценка разных методов передачи данных
- ❑ Возможные типовые операции передачи данных
- ❑ Полезность использования логических топологий
- ❑ Достаточность рассмотренного множества типовых операций передачи





# Литература

---

- ❑ **Гергель В.П.** Теория и практика параллельных вычислений. – М.: Интернет-Университет, БИНОМ. Лаборатория знаний, 2007.
- ❑ **Kumar V., Grama, A., Gupta, A., Karypis, G. (1994).** Introduction to Parallel Computing. - The Benjamin/Cummings Publishing Company, Inc. (2nd edn., 2003)
- ❑ **Andrews, G. R. (2000).** Foundations of Multithreaded, Parallel, and Distributed Programming.. – Reading, MA: Addison-Wesley (русский перевод Эндрюс Г.Р. Основы многопоточного, параллельного и распределенного программирования. – М.: Издательский дом "Вильямс", 2003).

