

Лекция 2.

**Архитектура параллельных
вычислительных систем. Часть 1.
История и проблематика**

Основы параллельного
программирования.

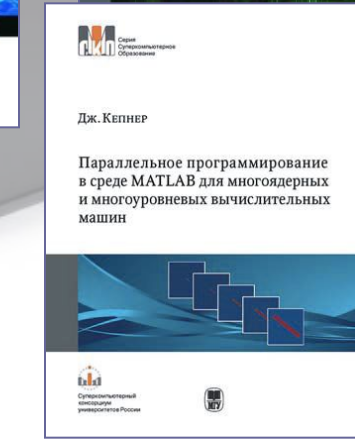
М.А. Сокольская



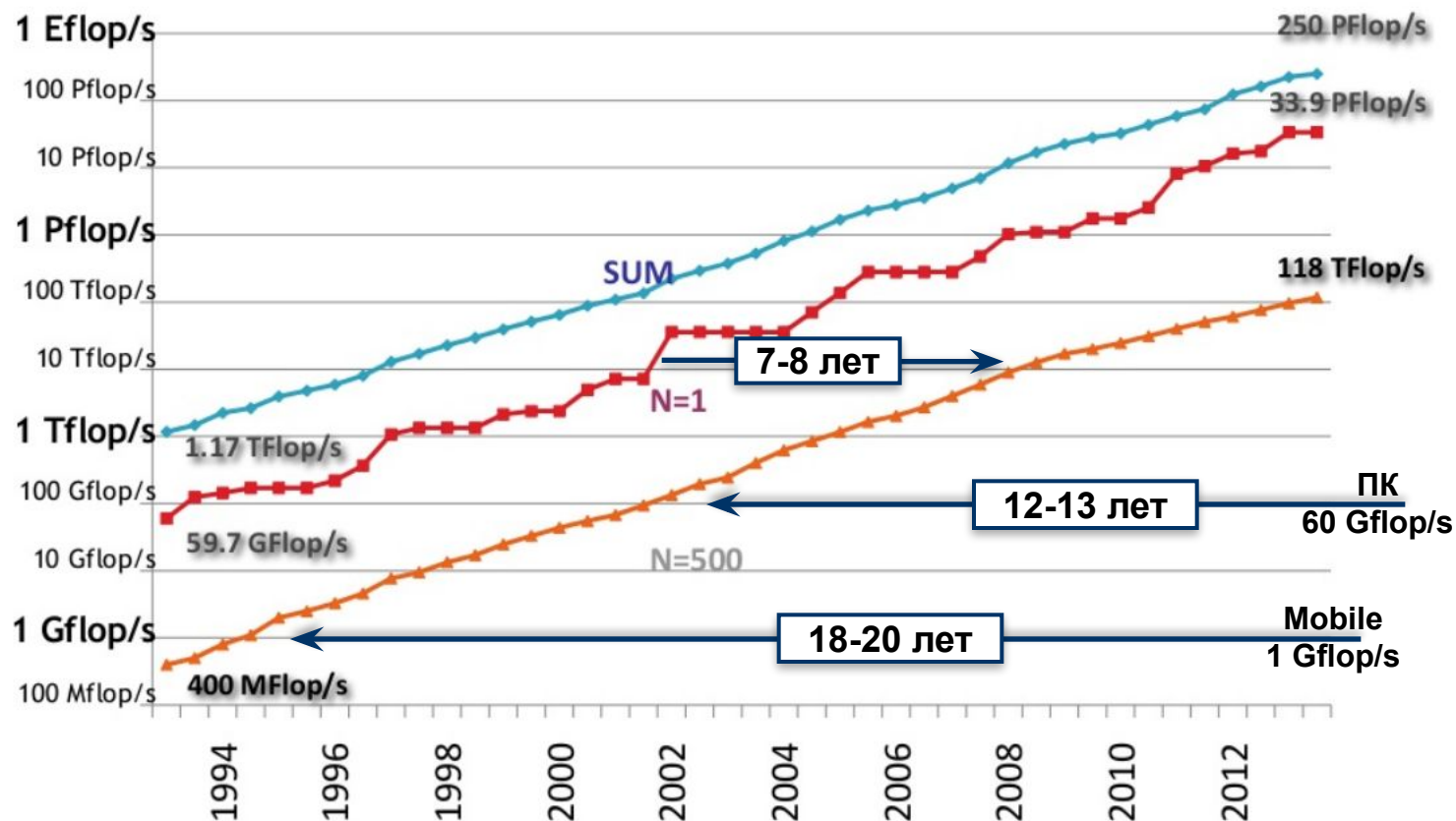
Литература

1. Гергель В.П. Теория и практика параллельных вычислений. - М.: Интернет-Университет, БИНОМ. Лаборатория знаний, 2007.
2. Богачев К.Ю. Основы параллельного программирования. - М.: БИНОМ. Лаборатория знаний, 2003.
3. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. - СПб.: БХВ-Петербург, 2002.
4. Немнюгин С., Стесик О. Параллельное программирование для многопроцессорных вычислительных систем — СПб.: БХВ-Петербург, 2002.

Серия книг “Суперкомпьютерное образование” – лауреат национальной премии “Книга года” (номинация “Учебник XXI века”)



Рост производительности вычислительных систем



Почему нужны параллельные вычисления...

- Опережение потребности в вычислениях быстрогодействия существующих компьютерных систем
- Оценка необходимой производительности – 10^{18} операций над вещественными числами в секунду (1 Eflops) – 1 квинтиллион операций в секунду.

Возможно, экзафлоп будет достигнут к 2020-2022 гг.

Почему нужны параллельные вычисления...

- Теоретическая ограниченность роста производительности последовательных компьютеров
- Резкое снижение стоимости многопроцессорных (параллельных) вычислительных систем
 - 1 Cray T90 processor – 1.8 GFlops (\$2 500 000),
 - 8 Node IBM SP2 using R6000 - 2.1 GFlops (\$500 000)
- Смена парадигмы построения высокопроизводительных процессоров - многоядерность

Сдерживающие факторы...

- **высокая стоимость параллельных систем** – в соответствии с *законом Гроша (Grosch)*, производительность компьютера возрастает пропорционально квадрату его стоимости
?!
- **потери производительности для организации параллелизма** – согласно *гипотезе Минского (Minsky)*, ускорение, достигаемое при использовании параллельной системы, пропорционально двоичному логарифму от числа процессоров
?!

Сдерживающие факторы...

- **постоянное совершенствование последовательных компьютеров** – в соответствии с *законом Мура (Moore)* мощность последовательных процессоров возрастает практически в два раза каждые 18 месяцев
?!
- **существование последовательных вычислений** – в соответствии с *законом Амдаля (Amdahl)* ускорение процесса вычислений при использовании p процессоров ограничивается величиной

$$S \leq 1/(f+(1-f)/p) \leq 1/f,$$

где f есть доля последовательных вычислений в применяемом алгоритме обработки данных
?!

Сдерживающие факторы...

- зависимость эффективности параллелизма от учета характерных свойств параллельных систем (отсутствие мобильности для параллельных программ)
?!

Пути достижения параллелизма...

- Достижение параллелизма возможно только при выполнимости следующих требований:
 - **независимость функционирования отдельных устройств ЭВМ** (устройства ввода-вывода, обрабатывающие процессоры и устройства памяти),
 - **избыточность элементов вычислительной системы**
 - *использование специализированных устройств* (например, отдельные процессоры для целочисленной и вещественной арифметики, устройства многоуровневой памяти),
 - *дублирование устройств ЭВМ* (например, использование нескольких одноптипных обрабатывающих процессоров или нескольких устройств оперативной памяти),
 - Дополнительная форма обеспечения параллелизма - **конвейерная** реализация обрабатывающих устройств

Пути достижения параллелизма...

- Возможные режимы выполнения независимых частей программы:
 - *многозадачный режим (режим разделения времени)*, при котором для выполнения нескольких процессов используется единственный процессор,
 - *параллельное выполнение*, когда в один и тот же момент времени может выполняться несколько команд обработки данных (обеспечивается при наличии нескольких процессоров или при помощи конвейерных и векторных обрабатывающих устройств),
 - *распределенные вычисления*, при которых для параллельной обработки данных используется несколько обрабатывающих устройств, достаточно удаленных друг от друга, а передача данных по линиям связи приводит к существенным временным задержкам.



*Основное внимание будем уделять второму
типу организации параллелизма,
реализуемому на многопроцессорных
вычислительных системах*

Примеры параллельных вычислительных систем...

- Суперкомпьютеры

Существует мировой рейтинг суперкомпьютерных систем: ТОП500.

<http://www.top500.org>

Рейтинг содержит 500 лучших машин мира, приводит множество их характеристик, позволяет оценить тенденции.

Производительность машин оценивается стандартным тестом LINPACK.

Исторические примеры

- **Суперкомпьютеры. Программа ASCI (США)**

(Accelerated Strategic Computing Initiative)

- **1996**, система **ASCI Red**, построенная Intel, производительность 1 TFlops,
- **1999**, **ASCI Blue Pacific** от IBM и **ASCI Blue Mountain** от SGI, производительность 3 TFlops,
- **2000**, **ASCI White** с пиковой производительностью свыше 12 TFlops (реально показанная производительность на тесте LINPACK составила на тот момент 4938 GFlops)

Примеры параллельных вычислительных систем...

- **Суперкомпьютеры. Система BlueGene**

- Первый вариант системы представлен в 2004 г. и сразу занял 1 позицию в списке Top500
- Расширенный вариант суперкомпьютера (ноябрь 2007 г.) по прежнему на 1 месте в перечне наиболее быстродействующих вычислительных систем
- **Сейчас – 22 место в рейтинге**
 - 212992 двухядерных 32-битных процессоров PowerPC 440 0.7 GHz,
 - пиковая производительность около 600 Tflops, производительность на тесте **LINPACK** – 478 Tflops

Примеры параллельных вычислительных систем...

- **Кластеры**

Кластер – группа компьютеров, объединенных в локальную вычислительную сеть (ЛВС) и способных работать в качестве единого вычислительного ресурса.

Предполагает более высокую надежность и эффективность, нежели ЛВС, и существенно более низкую стоимость в сравнении с другими типами параллельных вычислительных систем (за счет использования типовых аппаратных и программных решений).

Примеры параллельных вычислительных систем...

- **Кластеры. Beowulf**

- В настоящее время под кластером типа “*Beowulf*” понимается вычислительная система, состоящая из одного серверного узла и одного или более клиентских узлов, соединенных при помощи сети Ethernet или некоторой другой сети передачи данных. Это система, построенная из готовых серийно выпускающихся промышленных компонент, на которых может работать ОС Linux/Windows, стандартных адаптеров Ethernet и коммутаторов.

Примеры параллельных вычислительных систем...

Кластеры. Beowulf...

- **1994**, научно-космический центр NASA Goddard Space Flight Center, руководители проекта - Томас Стерлинг и Дон Бекер:
 - 16 компьютеров на базе процессоров 486DX4, тактовая частота 100 MHz,
 - 16 Mb оперативной памяти на каждом узле,
 - Три параллельно работающих 10Mbit/s сетевых адаптера,
 - Операционная система Linux, компилятор GNU, поддержка параллельных программ на основе MPI.

Примеры параллельных вычислительных систем...

- **Кластеры. AC3 Velocity Cluster**

- **2000**, Корнельский университет (США), результат совместной работы университета и Advanced Cluster Computing Consortium, образованного компаниями Dell, Intel, Microsoft, Giganet:
 - 64 четырехпроцессорных сервера Dell PowerEdge 6350 на базе Intel Pentium III Xeon 500 MHz, 4 GB RAM, 54 GB HDD, 100 Mbit Ethernet card,
 - 1 восьмипроцессорный сервер Dell PowerEdge 6350 на базе Intel Pentium III Xeon 550 MHz, 8 GB RAM, 36 GB HDD, 100 Mbit Ethernet card,
 - Операционная система Microsoft Windows NT 4.0 Server Enterprise Edition,
 - Пиковая производительность AC3 Velocity 122 GFlops, производительность на тесте LINPACK 47 GFlops.

• Кластеры. Thunder

– **2004**, Ливерморская Национальная Лаборатория (США):

- 1024 сервера, в каждом по 4 процессора Intel Itanium 1.4 GHz,
- 8 Gb оперативной памяти на сервер,
- общая емкость дисковой системы 150 Tb,
- операционная система CHAOS 2.0,
- пиковая производительность 22938 GFlops и максимально показанная на тесте LINPACK 19940 GFlops (5-ая позиция списка Top500).

Примеры параллельных вычислительных систем...

- **Кластер. Ломоносов-2**

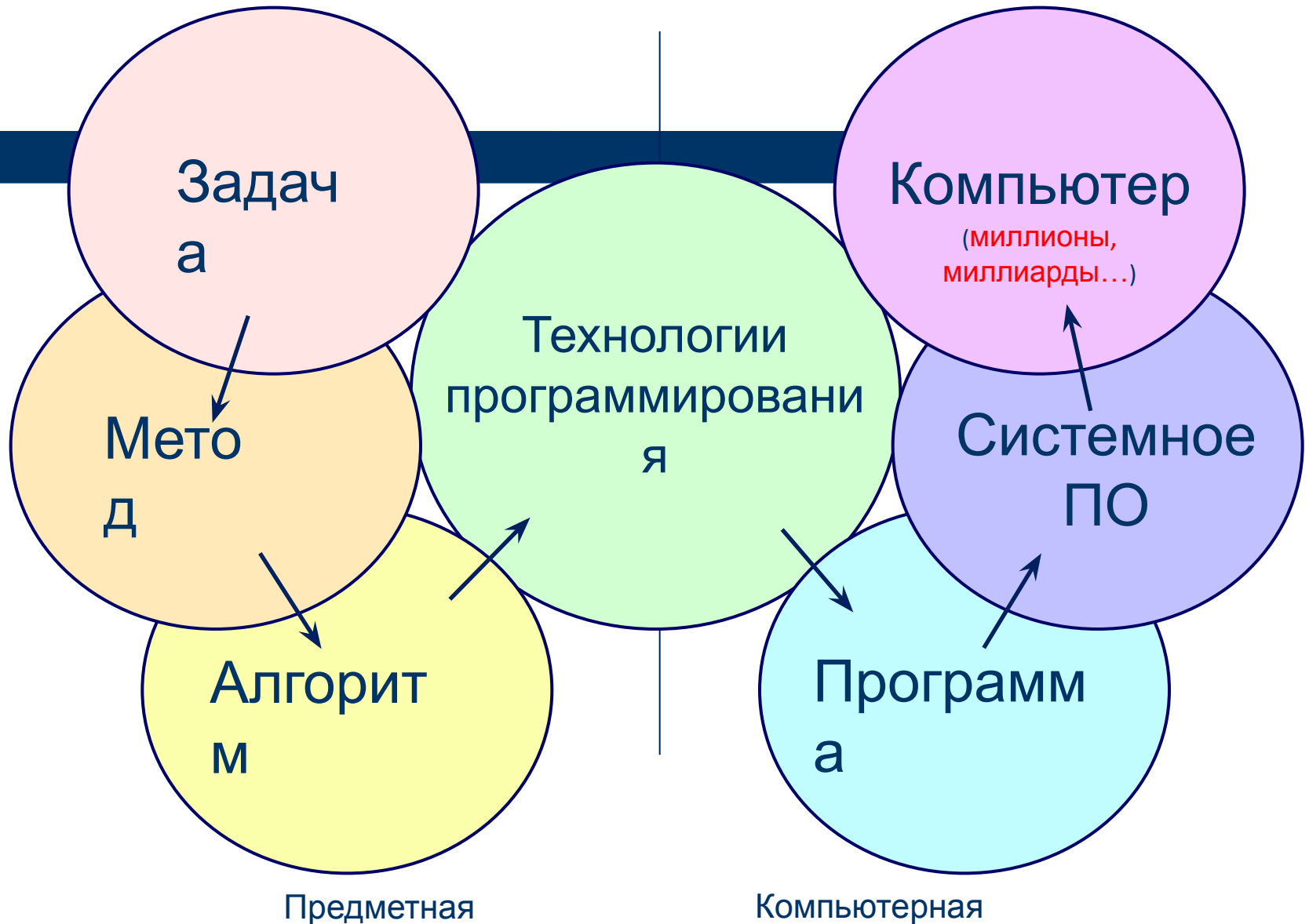
(МГУ им. М.В. Ломоносова)

2016, 41 место в рейтинге на июнь (недостроенная система)

- Пиковая производительность — 1,7 Рфлопс
- Производительность на тесте LINPACK – 900 Tflops
 - Число процессорных ядер – 78660, из них 29820 на ускорителях
 - Оперативная память — 99489 ГБ
 - Объем системы хранения данных около 2000 ТБ
 - операционная система — Clustro операционная система — Clustrix T-Platforms Edition
 - Занимаемая площадь — 252 кв.м

Выполнение задачи на компьютере

(от мобильных платформ до экзафлопсных суперкомпьютерных систем)



Суперкомпьютерные центры: ожидания и реальность

1 Pflop/s system... Что мы ожидаем?

$1 \text{ Pflop} * 60 \text{ sec} * 60 \text{ min} * 24 \text{ hours} * 365 \text{ days} = 31,5 \text{ ZettaFlop } (10^{21}) \text{ per year}$

Что в реальности? Очень малая доля...

Суперкомпьютеры и паровозы...
Где эффективность выше?



Особенности архитектуры, сложные потоки задач, плохая локальность данных, огромная степень параллелизма в аппаратуре и т.п. – все это причины низкой эффективности.

Эффективность суперкомпьютерных центров (простая оценка)



Усредненная производительность одного ядра
суперкомпьютера “Чебышев” за 3 дня

Эффективность суперкомпьютерных центров

Проекты, приложения

Очереди заданий

Динамика приложения

Поток приложений

Стек ПО

Компоненты

компьютера

Инж.инфраструктура

*CPU usage:
user, system, irq, io, idle,
(summary, and per-core)
Performance counters;
Swap usage;
Memory usage;
Interconnect usage;
Network errors;
Disk usage;
Filesystem usage;
Network filesystem usage;
Hardware alarms (ECC, SMART, etc);
CPU and motherboard temperatures;
Network switches errors;
Cooling subsystem data;
Power subsystem data;
FAN speeds;
Voltages;
...*

Крайне сложная структура суперкомпьютеров привела к потере контроля над пониманием (знанием) их поведения.

Цель - полный контроль над HW/SW и приложениями.

Большие числа суперкомпьютерных центров (степень параллелизма)

Сегодня в суперкомпьютерах все параметры большие :

- Ядра, процессоры, ускорители, узлы,
- Компоненты аппаратуры,
- Компоненты программного обеспечения,
- Число файлов в СХД, число буферов,
- Трафик в интерконнекте,
- Число пользователей, проектов,
- Выполняющиеся и ожидающие задачи,
- ...

И все эти числа очень быстро растут !