



# **ИПС имени А.К. Айламазяна РАН — 30 лет. Суперкомпьютерные разработки**

Абрамов С.М.

Национальный суперкомпьютерный форум  
2014.11.25, Переславль-Залесский





## Создание Института. Изначальные задачи

- **23.02.1984** постановлением Президиума АН СССР № 226 создан как филиал Института проблем кибернетики АН СССР
- Директор: д.т.н., профессор А.К. Айламазян

**Комплекс для НРС:** градирни, бассейн, ЛЭП, три подстанции, машинный зал 800 м<sup>2</sup>, без ограничения нагрузки на пол, h=9м





# 1984–1992: Только отечественные суперкомпьютеры





## Отечественные НРС-технологии

- **Предыстория:** работа в НИЦЭВТ...
  - ☆ **Спецпроцессоры ЕС ЭВМ:** ЕС 2700, ЕС 2701, ЕС 2703, ЕС 2704, ЕС 2706...
  - ☆ **Старшие машины ЕС ЭВМ:** ЕС 1066, ЕС 1090
  - ☆ **Вне ЕС ЭВМ:** АСВТ, Эльбрус-1/2, ПС-1000/ПС-2000, Электроника СС БИС, НИИ «Квант»
- **1984–1992:** участие ИПС РАН в работах **МДА ЕС 2704** — 6 интерфейсных, 12 коммутационных и 24 вычислительных модуля <http://goo.gl/2wikgT>
  - ☆ НИЦЭВТ, ЛИИ АН СССР
  - ☆ Гл. идеолог проф. В.А.Торгашев, Гл. конструктор В.У. Плюснин, Гл. разработчик ПО И.Р. Агамирзян
  - ☆ Ассемблер МДА ЕС 2704; язык Си; задача «Демультимплексор»



# **1990–1995: Начало эпохи импортом-замещения. Транспьютерные системы**





## 1990–1995: Начало импортом-замещения. Российская транспьютерная ассоциация

**1990–1995:** работы с транспьютерными системами, участие Института в Российской транспьютерной ассоциации

### □ Партнеры:

- ☆ НИИ «Квант», проф. А.И.Галушкин
- ☆ Международная лаборатория ИИ в Братиславе
- ☆ Университет Катания и лаборатория Inmos (Сицилия, Италия) — грант Intas
- ☆ High Performance Computing Laboratory (Афины, Греция)

□ **TTOLS, компилятор gcc для T4, T8, T9** (Шевчук Ю.В.), победили ACE из ОС Parix (Parsytec, Gcel) <http://goo.gl/h28kBU>

□ **Интерфейсная плата DAD005** (Пономарёв А.Ю., Шевчук Ю.В., Позлевич Р.В.) на базе T425: ISA, разделяемая память, 5 Мбит/с — на порядок быстрее, чем через интерфейсный чип C011

□ **Алгоритмы** динамической балансировки нагрузки и маршрутизации (по сути OSPF) в произвольных транспьютерных сетях

□ **Исследования топологий** интерконнектов: лаборатория ЗМР, графы минимального диаметра, теоретические работы и инструментальное ПО



**1995–1998:**  
**Начало проекта «Т-система».**  
**Первые кластеры, «потешные» и**  
**не очень...**





## Начало проекта «Т-система». Первые кластеры, «потешные» и не очень...

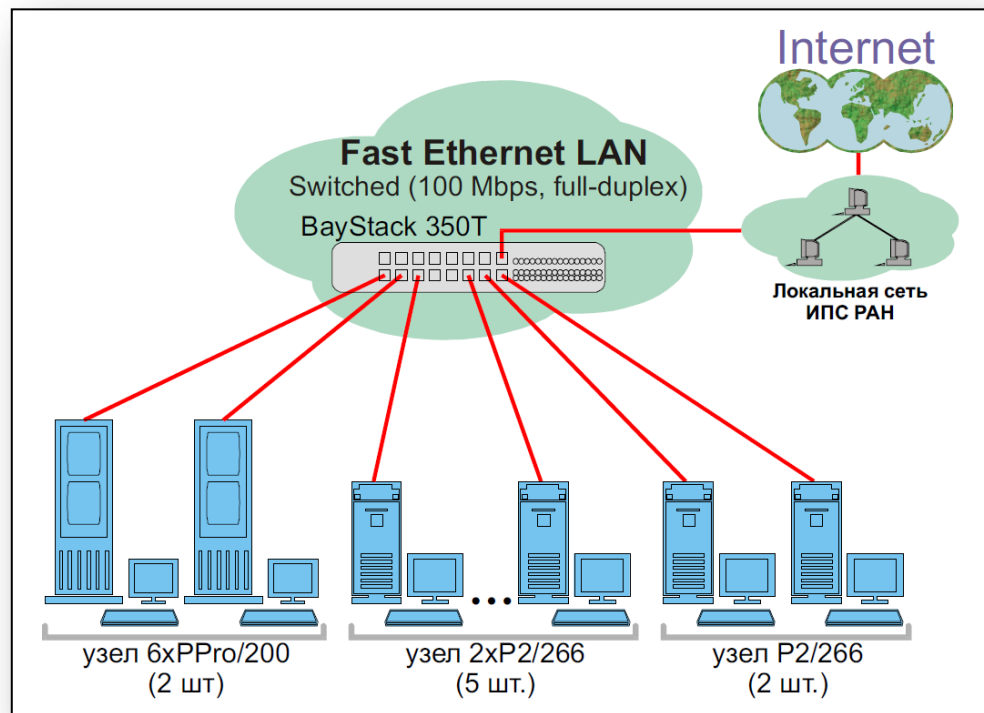
- **1995–1998:** поиск и реализация решений для компонент первых версий Т-системы
  - ☆ Функциональные диалекты императивных языков
  - ☆ Представление процесса вычисления в виде автотрансформации вычисляемой сети
  - ☆ Автоматическое динамическое распараллеливание программ
  - ☆ **Весна 1998: первое Т-ядро** (списковые структуры, «далекие указатели», пересылка значений, отношение «поставщик—потребитель», распределенная сборка мусора, отладка вдоль трассы), **первая Т-задача** — трассировка лучей
- **1995–1998:** поиск и реализация платформ для наших разработок
  - ☆ OS Linux
  - ☆ **Сети из ПЭВМ** соединенных ускоренным RS-232 (до 1 Mbit/s) и собственными коммутирующими устройствами для него (коммутация каналов)
  - ☆ **Сети из ПЭВМ** на основе 10 Base-2
  - ☆ **1998:** первый серьезный кластер на базе FastEthernet (100 Mbit/s)





## 1998: Программно-аппаратный мультипроцессорный комплекс 5.6 Gflops

- ❑ **Весна 1998:** Первые SMP-компьютеры в Институте, первый Fast Ethernet
- ❑ 2×6 Intel PPro-200+5×2 Intel P-II-266+2×1 Intel P-II-266, 5.6 Gflops (пик), RAM 1.4 GB, HDD 70.4 GB
- ❑ Поддержка SMP в T-системе
- ❑ (Псевдо-)компилятор для T-системы (на Рефале)
- ❑ **За полгода: 10 различных T-задач**





**1998–1999:  
Рождение СКИФа.  
Россия и Беларусь: частичное  
восстановление расколотого  
МРП СССР**





## Начало СКИФ-кооперации

- **Май 1998:** визит в Минск к Президенту НАН Беларуси (Александр Павлович Войтович)
  - ☆ **Т-система:** С.М. Абрамов, М.Р. Коваленко
  - ☆ **ОВС:** В.Ю. Татур, компания «Суперкомпьютерные системы» СКС
- **Июнь 5–12 1998:** ответный визит, «Договор о неравном браке»
  - ☆ Доклад А.П. Войтовича Президенту Республики Беларусь
  - ☆ Руководящие решения
- **1998:** временный трудовой коллектив с коллегами из Минска, формирование суперкомпьютерной программы «СКИФ» Союзного государства
- **Весь 1999:** согласование в Министерствах
- **С сентября 2000:** начало работ Программы «СКИФ»
- **1998–1999:** развитие первой версии Т-системы



# **2000–2004: Программа «СКИФ» Суперкомпьютеры «СКИФ» ряд 1 и 2**





## 2000–2004: Программа «СКИФ» Суперкомпьютеры «СКИФ» ряд 1 и 2

- ❑ **10 организаций от Беларуси и 10 организаций от России:** ИПС имени А.К. Айламазяна РАН (головной от России), НИЦЭВТ, НИИ КС, ЦНТК РАН, СКС, РосНИИ РП, ИВВиБД (С-Пб), НИИ механики МГУ имени М.В. Ломоносова и др.
- ❑ **17 опытных образцов** систем Ряда 1 и Ряда 2 семейства «СКИФ», КД и ПД ряда 1 и 2 (литера  $O_1$ )
- ❑ Экспериментальный гибридный макет «**ОВС+СКИФ Первенец**»
- ❑ Производство SCI адаптеров **N330, N337, N335** (НИЦЭВТ)
- ❑ **Развитие T-Системы:**
  - ☆ GRACE (1999–2002) и OpenTS (с 2003)
  - ☆ Синтаксически-гладкое расширение C++, «настоящий» компилятор, качественная RS-система
  - ☆ T++, T-Fortran



## 2000–2004: Программа «СКИФ» Суперкомпьютеры «СКИФ» ряд 1 и 2

- ❑ **Сервисные сети** СКИФ ServNet v.1 и ServNet v.2
  - ☆ Селективный on/off, Селективный reset, Linux serial console, «черный ящик»
- ❑ Ядро операционной системы **Linux-SKIF** (ИПС РАН и МГУ)
- ❑ **Отладчик TDB** — импортозамещение TotalView
- ❑ **PVFS-SKIF, OpenPBS-SKIF**, мониторинговая система **FLAME-SKIF** (ИПС РАН и МГУ)
- ❑ **Первые ≈20 прикладных систем**, в том числе:
  - ☆ автоматизация проектирования химических реакторов (ИВВиИС, СПб.)
  - ☆ классификация потока текстов (ИЦИИ ИПС РАН)
  - ☆ **Т-приложения:** «MultiGen», «вывоз мусора (Бельгия)», «обтекание плохо обтекаемых тел», «ДЗЗ», «радиолокация»



# Суперкомпьютеры «СКИФ» Ряд 1



2000: СКИФ Первенец



2001: СКИФ VM-5100



2003 — СКИФ  
ES1710.03 ТКС



## Суперкомпьютеры «СКИФ» Ряд 2

2003: СКИФ T-Forge-32



2003: СКИФ К-500 Top#405 11.2003



2004, СКИФ T-Forge20NB



2004, ускоритель Однородная ВС



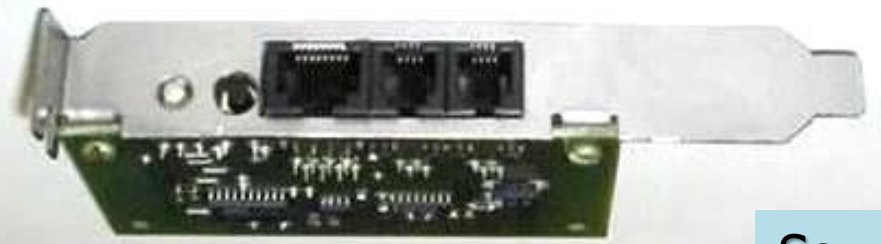
2004: СКИФ К-1000 11.2004 Top#98



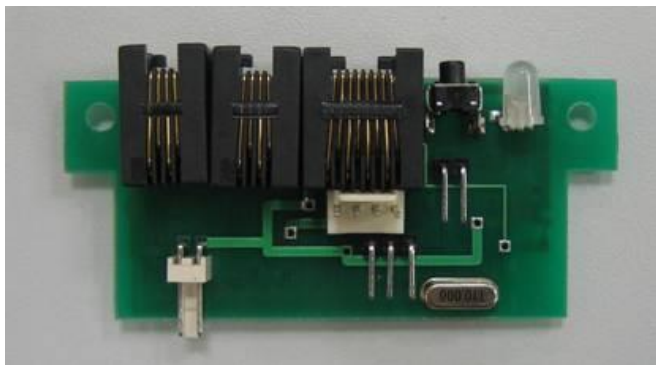
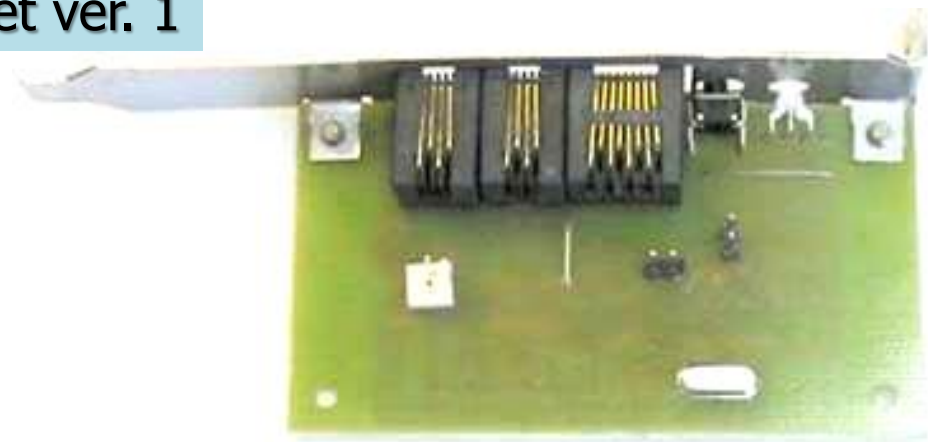
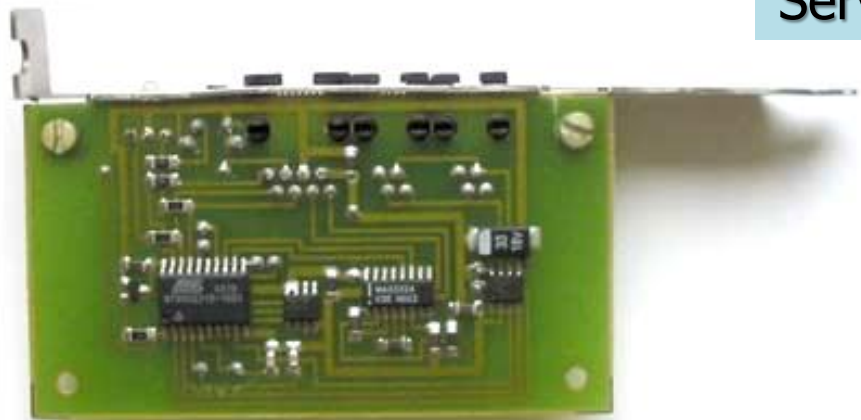




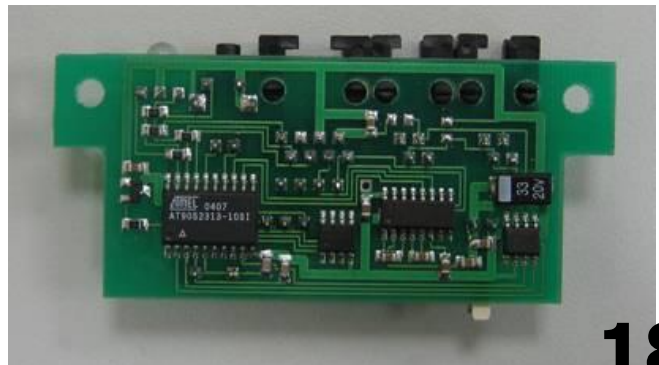
# Сервисная сеть для суперЭВМ «СКИФ» Ряда 2 Servnet ver.1 и ver.2



Servnet ver. 1



Servnet ver. 2





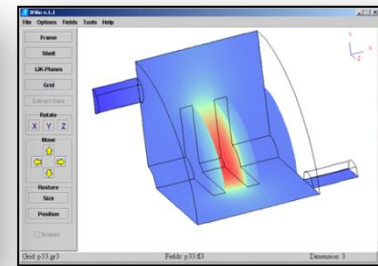
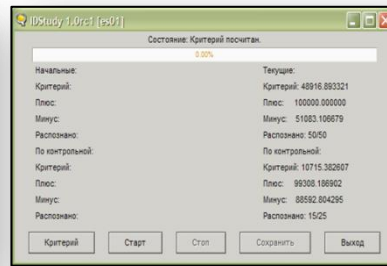
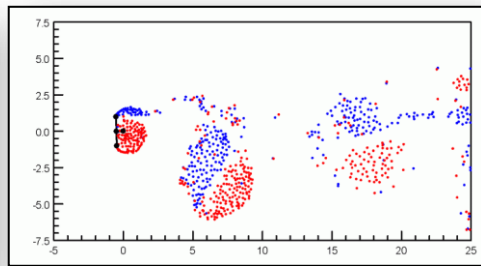
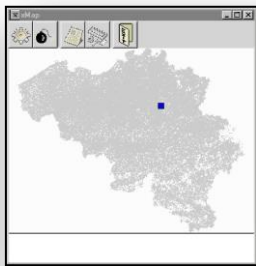
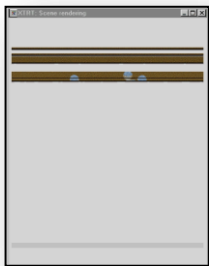
## Комплект ПО суперкомпьютеров «СКИФ»



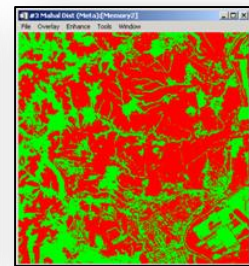
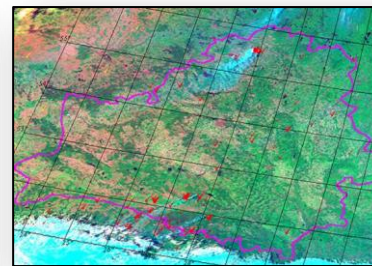
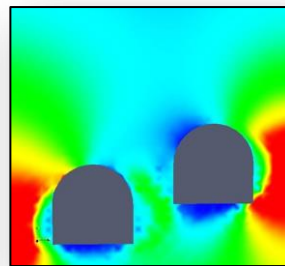
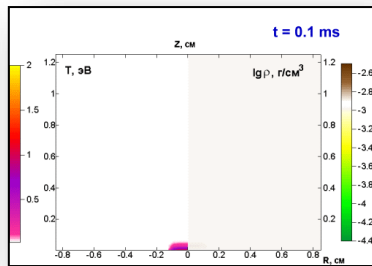
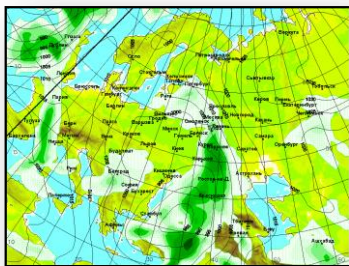
- ❑ **системное ПО:** операционная система; базовые библиотеки поддержки параллельного счета; файловые системы; системы очередей, мониторинга и управления; стандартные системы программирования — С, С++, Fortran; и т. п.
- ❑ **средства разработки параллельных программ** — программные системы, инструментальные средства и библиотеки: Grace, Open TS, MIRACLE и др.;
- ❑ **два десятка параллельных прикладных систем** для различных областей.



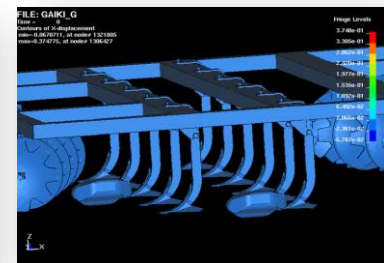
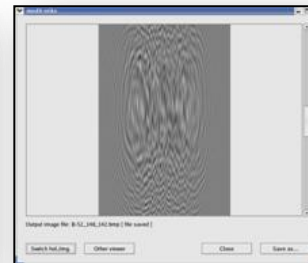
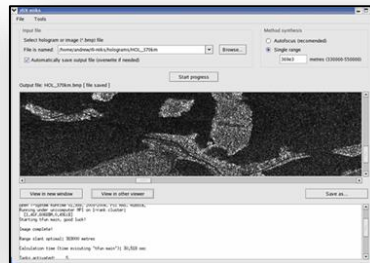
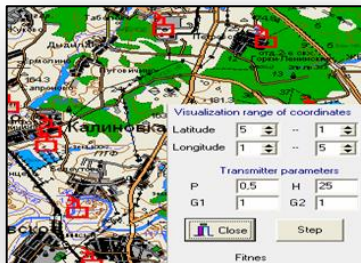
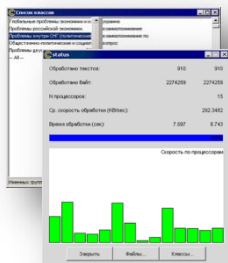
# 2000–2004: Программа «СКИФ» Первые 20 отечественных приложений



рендеринг, транспортная задача, обтекание плохообтекаемых тел, распознавание по голосу, проектирование химреакторов



модель Лосева, факел лазера, прочность шахт, расчет загрязнений от лесных пожаров, классификация по ДЗЗ



классификация текстов, распределение РЭС на поле боя, перспективные катализаторы, совместимость с LS-DYNA



## Признание государствами

### Новогоднее поздравление Президента Республики Беларусь А.Г. Лукашенко белорусскому народу (31 декабря 2003 г.)



- «Время — птица» ... хрупок мир ... государства, как и люди, должны следовать нравственным принципам ...
- **Мы достойно прожили уходящий год. Нам есть чем гордиться**
  - ☆ **Создав суперкомпьютер «СКИФ», наша страна вошла в элиту мирового информационного сообщества**
  - ☆ Беларусь входит в список космических держав мира. В ближайшее время у нас будет свой искусственный спутник.
  - ☆ Мы завершили газификацию Полесья...
  - ☆ Построены десятки тысяч домов...
  - ☆ Немало сделано по наведению порядка на нашей земле...
  - ☆ ... собран такой урожай, что мы можем поделиться с другими ...
  - ☆ Год уходящий отмечен яркими событиями в духовной сфере, культуре и спорте ...



## **2005–2007: Бюрократические провоочки и потеря темпа.**

**Инициатива «СКИФ — Университетам»  
и СКИФ Cyberia**

**OpenTS и Microsoft**





## 2005–2007: Бюрократические проволочки и потеря темпа

- ❑ **2004–март 2007:** три с половиной года согласования «СКИФ-ГРИД»

### Инициативные разработки

- ❑ **Программа «СКИФ — Университетам»:**

- ☆ ИПС имени А.К. Айламазяна РАН, Т-Платформы, «Единая Россия»
- ☆ СКИФ-Cyberia

- ❑ **Переславский T-Grid:** до 120 CPU в системе

- ❑ **2006–2007: библиотека T-Sim** параллельного программирования на основе шаблонных классов C++

- ❑ **2005–2007: распределенная система мониторинга** технических параметров оборудования в нескольких суперкомпьютерных центрах

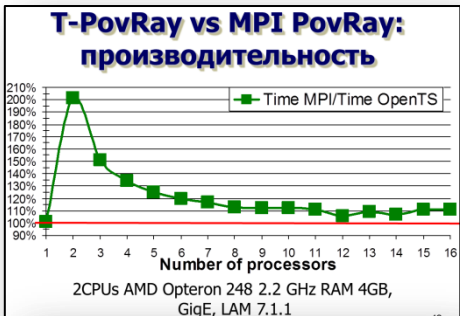
- ❑ **2006: Web-сервисы на базе T-системы**

- ❑ **2006–2007: Задел для СКИФ-ServNet v.3** — поддержка blade-решений и десятков тысяч узлов

- ☆ **Будет востребована в 2008 году:** 800 плат ServNet T-60 и 80 плат ServNet CNB для СКИФ МГУ (Чебышёв) и СКИФ Урал

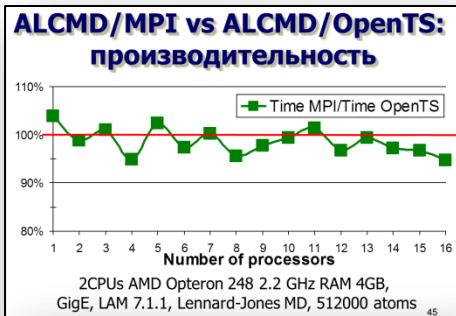


# 2005–2006: OpenTS и Microsoft. SC'06 Tampa, Microsoft Radmond, USA



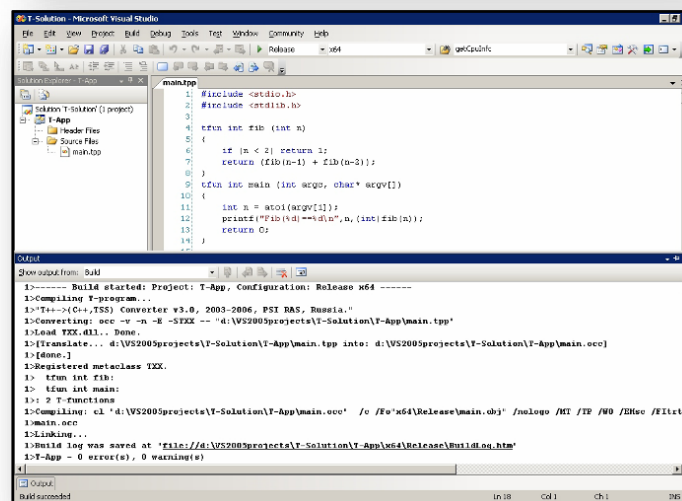
Программа	Объем кода
MPI modules for PovRay 3.10g	1,500 строк
MPI patch for PovRay 3.50c	3,000 строк
T++ modules (for both versions 3.10g & 3.50c)	200 строк

~7–15 раз



Программа	Объем кода
MP_Lite total/MPI	~20,000 строк
MP_Lite,ALCMD-related/MPI	~3,500 строк
MP_Lite,ALCMD-related/OpenTS	500 строк

~7 раз



Кирилл Фаенов, Игорь Агамирзян,  
Бартон Смит и многие другие...



**2007–2010:  
Программа «СКИФ-ГРИД»  
Союзного государства**







## 2007–2010: Программа «СКИФ-ГРИД» Союзного государства

### Направления работ

- 1. GRID-технологии**, поддержка гетерогенных, территориально-распределенных вычислительных комплексов
- 2. Суперкомпьютеры семейства «СКИФ» (Ряд 3 и 4)**, поддержка перспективных процессоров, новых решений по интерконнекту, управления системой, спецвычислителей и гибридных узлов, разработка ПО
- 3. Защита информации**
- 4. Пилотные системы:** реализация прикладных систем в перспективных областях применения суперкомпьютеров и GRID-систем, усилия по подготовке и переподготовке кадров



## Программа «СКИФ-ГРИД». Второй этап (2009–2010). Российские участники — 37 организаций

### Учреждения РАН

#### 1. **Головной от России:**

- ИПС имени А.К. Айламазяна РАН
- ИММ РАН
- ГЦ РАН
- ИКИ РАН
- ИСА РАН
- ИПМ имени М.В. Келдыша РАН
- ИППИ РАН
- ИБХФ РАН
- ИПХФ РАН
- ИХФ РАН

### ВУЗы и НИИ ВУЗов

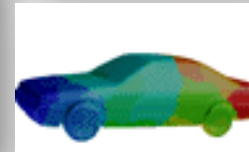
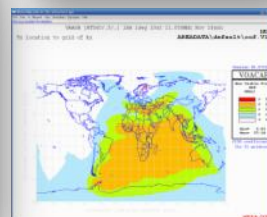
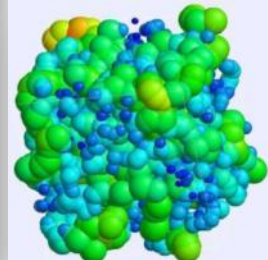
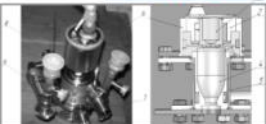
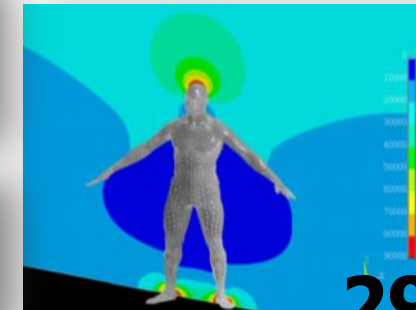
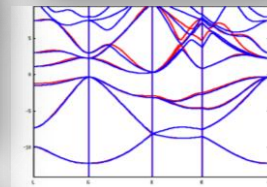
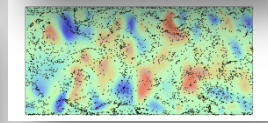
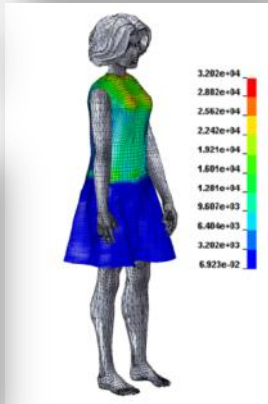
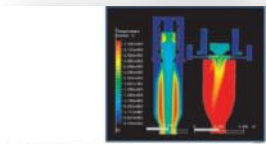
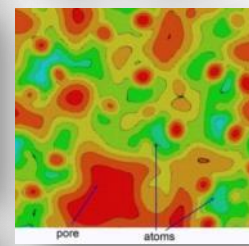
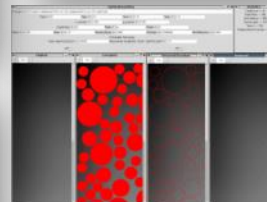
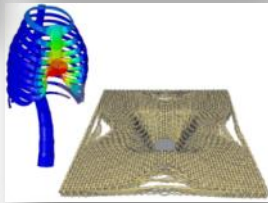
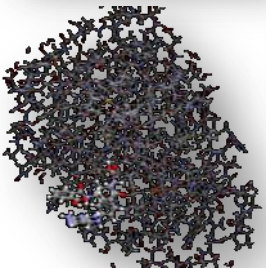
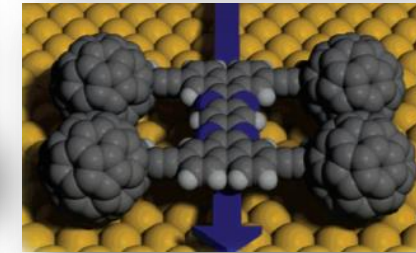
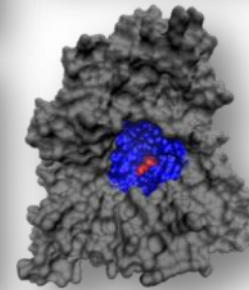
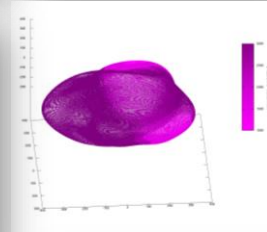
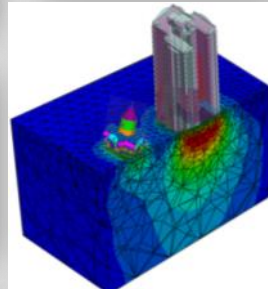
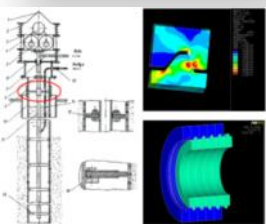
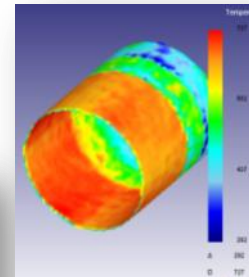
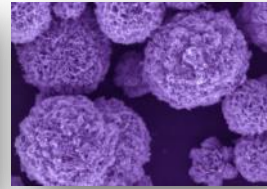
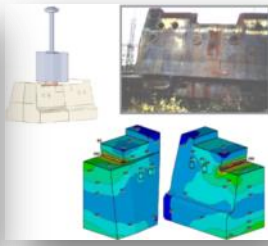
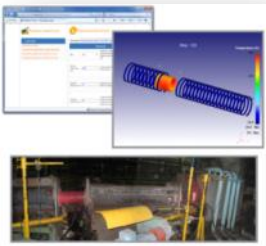
- ЮУрГУ
- УГАТУ
- МТУСИ
- ННГУ
- СПбГПУ
- ТГУ
- ВлГУ
- ПензГУ
- ЧелГУ
- МГУ имени М.В. Ломоносова:
- ВМК МГУ
- НИИЯФ МГУ
- ХФ МГУ
- НИВЦ МГУ
- НИИФХБ МГУ

### НИИ, предприятия наукоемких отраслей

- НИИ КС
- СПБАЭП
- ЦНИИ МАШ
- НПЦ «Элвис»
- «Каледин и Партнеры»
- «РСК СКИФ»
- «Альт Линукс Технолоджи»
- «НИЦЭВТ»
- «ЮникАйСиз»
- «Сигма Технология»
- «Тесис»
- «Урал-Грид»
- «Кинтех»



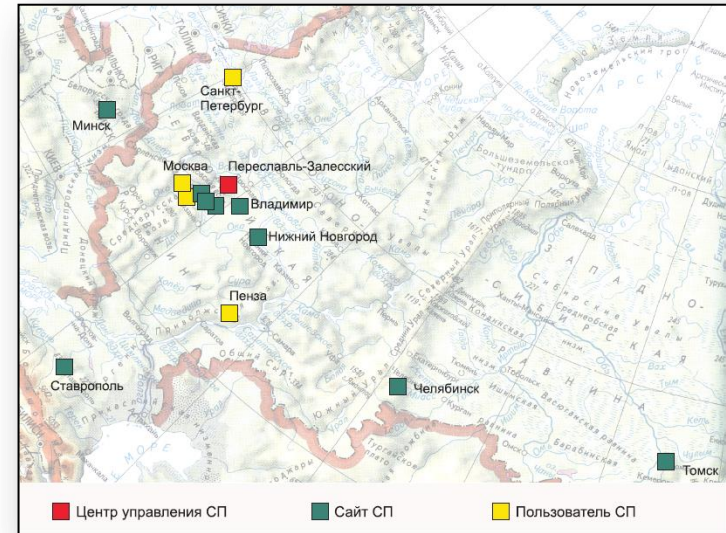
# Прикладные грид-сервисы и пилотные прикладные системы





## 2007–2010: СКИФ-ГРИД Некоторые разработки

- ❑ **SKIF@Home** (импортозамещение для VOINC)
- ❑ **СПОКУ**: система предсказания отказов кластерных установок, машинное обучение, big data
- ❑ **Развитие OpenTS**: поддержка многоядерности процессоров, поддержка grid-среды
- ❑ **ПО промежуточного уровня и СКИФ-Полигон**
- ❑ **«СКИФ МГУ» (Чебышёв)**  
**06.2008 Top#36**
- ❑ **«СКИФ Урал»**  
**06.2008 Top#282**












# Опытный образец «СКИФ Аврора ЮУрГУ»





## Преимущества СКИФ ряда 4








6x   
5x   
3x 

-  **все печатные платы, всю «механику», систему в целом можно изготавливать в России** (имеем право и способны), закупать надо только микросхемы (ЭКБ)
-  **в 1,5 раза лучше** эффективность использования электроэнергии
-  **в 2 раза плотнее** упаковка вычислительной мощности;
-  **в 1,5 раза выше** пропускная способность системной сети
-  **улучшенная масштабируемость** системной сети (3D-top) — преодоление ограничений на поставку
-  **повышенная эффективность** реализации синхронизации — преодоление ограничений на поставку
-  **повышенная эффективность** реализации массовых операций



## Преимущества СКИФ ряда 4

6x   
5x   
3x 

-  **возможность поддержки не только МРІ**, но и новых перспективных подходов к реализации параллельных вычислений — преодоление запретов на поставку
-  **высокая совместимость** с существующим программным обеспечением **в комбинации с возможностью**
-  **использование FPGA-ускорителей** — преодоление ограничений на поставку
-  **повышенная надежность** суперкомпьютера (нет подвижных частей, N+1 резервирование, тройное резервирование управления и мониторинга)
-  **улучшенная система электропитания**
-  **улучшенные эргономические и эстетические показатели** — сенсорный мультитэкран для управления, вся система — беззвучная
-  **широкая отечественная кооперация:**
  - ☆ 7 организаций — разработка КД и ПД СКИФ-Аврора
  - ☆ 20 российских организаций — адаптация и оптимизация приложений



# **Российская НРС-отрасль и «СКИФ» + «СКИФ-ГРИД»: многое вышло из этой шинели**







## Суперкомпьютерные программы «СКИФ» и «СКИФ-ГРИД» Союзного государства

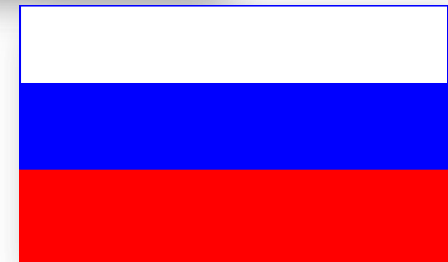
### □ Заказчики-координаторы

- ☆ Министерство образования и науки
  - ранее — агентство «Роснаука»
- ☆ НАН Беларуси



### □ Главные исполнители

- ☆ Институт программных систем РАН
- ☆ Объединенный институт проблем информатики НАН Беларуси



### □ «СКИФ» — 2000–2004

из бюджета России: 125 млн.руб.

### □ «СКИФ-ГРИД» — 2007–2010

из бюджета России: 446 млн.руб.

**5+4 = 9 лет**  
**571 млн. руб.**

### □ Исполнители (Россия+Беларусь)

- ☆ «СКИФ» — более: 10 (РФ) + 10 (РБ) организаций
- ☆ «СКИФ-ГРИД» — более: 37 (РФ) + 10 (РБ) организаций



# Суперкомпьютеры «СКИФ» в Top500

## 6 систем 22 раза вошли в рейтинг Top500






















# «СКИФ» и «СКИФ-ГРИД»: в России многое вышло из этой шинели...

No	Дата	Tflops	Суперкомпьютер	Разработчик
1	2002/06	0.734	MBC 1000M	ФГУП «НИИ Квант», МСЦ РАН, ИПМ имени М.В.Келдыша РАН
2	2003/11	0.423	СКИФ К-500	СКИФ-кооперация
3	2004/11	2.03	СКИФ К-1000	СКИФ-кооперация
4	2007/06	9.01	СКИФ Cyberia	СКИФ-кооперация
5	2008/06	12.2	СКИФ Урал	СКИФ-кооперация
6	2008/06	47.1	СКИФ МГУ (Чебышев)	СКИФ-кооперация
7	2009/11	21.8	СКИФ-Аврора ЮУрГУ	СКИФ-кооперация
8	2009/11	350	Ломоносов	гр. комп. «Т-Платформы»
9	2012/11	376	Tornado МСЦ РАН	гр. комп. «РСК»
10	2012/11	147	Tornado ЮУрГУ	гр. комп. «РСК»
11	2014/06	320	Ломоносов-2 (A class)	гр. комп. «Т-Платформы»
12	2014/11	658	Tornado СПбГПУ	гр. комп. «РСК»
13	2014/11	289	GPU Blade Cluster ИИГУ	Niagara Computers, Supermicro
14	2014/11	170	PetaStream СПбГПУ	гр. комп. «РСК»



# «СКИФ» и «СКИФ-ГРИД»: в России многое вышло из этой шинели...

No	Дата	Tflops	Суперкомпьютер	Разработчик
1	2002/06	0.734	МВС 1000М	ФГУП «НИИ Квант», МСЦ РАН, ИПМ имени М.В.Келдыша РАН
2	2003/11	0.423	СКИФ К-500 	СКИФ-кооперация
3	2004/11	2.03	СКИФ К-1000 	СКИФ-кооперация
4	2007/06	9.01	СКИФ Cyberia 	СКИФ-кооперация
5	2008/06	12.2	СКИФ Урал 	СКИФ-кооперация
6	2008/06	47.1	СКИФ МГУ (Чебышев) 	СКИФ-кооперация
7	2009/11	21.8	СКИФ-Аврора ЮУрГУ 	СКИФ-кооперация
8	2009/11	350	Ломоносов 	 гр. комп. «Т-Платформы»
9	2012/11	376	Tornado МСЦ РАН 	 гр. комп. «РСК»
10	2012/11	147	Tornado ЮУрГУ 	 гр. комп. «РСК»
11	2014/06	320	Ломоносов-2 (A class)	 гр. комп. «Т-Платформы»
12	2014/11	658	Tornado СПбГПУ 	 гр. комп. «РСК»
13	2014/11	289	GPU Blade Cluster ННГУ	Niagara Computers, Supermicro
14	2014/11	170	PetaStream СПбГПУ 	 гр. комп. «РСК»



# **2010–2014: Национальные особенности перспективных подсистем охлаждения**



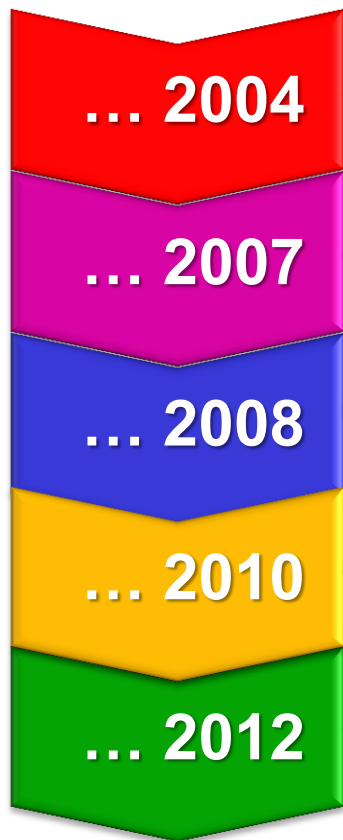


## Подсистема охлаждения суперЭВМ. Инженерная логика

- ☆ **Нужна всё большая производительность** →
- ☆ Нужны меньшие задержки →
- ☆ Нужны более короткие проводники →
- ☆ Нужна более плотная упаковка электроники →
- ☆ Будет более высокая объемная плотность тепловыделения →
- ☆ **Нужна всё более эффективная система охлаждения электроники**



# ИПС имени А.К. Айламазяна РАН: Эпохи развития технологий охлаждения



## Воздушное охлаждение

- неупорядоченное движение воздуха
- организованное движение воздуха
- многоконтурные системы «воздух—вода—фреон»
  - ☆ «жидкость на уровне ЦоД» — внутрирядные кондиционеры...
  - ☆ «жидкость на уровне шкафа» — холодные стенки...

## Жидкостное охлаждение

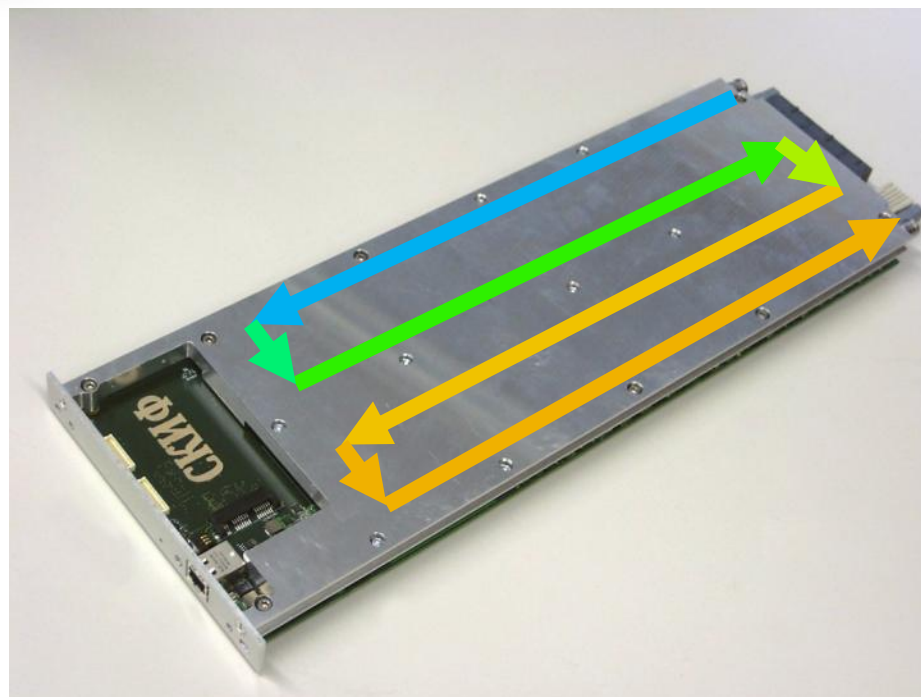
- закрытые системы жидкостного охлаждения
  - ☆ гибридное охлаждение:  $\frac{1}{2}$  — вода,  $\frac{1}{2}$  — воздух
- непосредственное (погружное) жидкостное охлаждение

## Что дальше?...

- Охлаждение фазовым переходом?...



# Закрытые системы жидкостного охлаждения на примере СКИФ-Аврора







## Типы систем **закрытого** жидкостного охлаждения

**Большие** охлаждающие пластины

**Гидравлический бэкплэйн,  
автоматическое подключение  
к охлаждению**

**Большие** охлаждающие пластины

Нет гидравлического бэкплэйна,  
**ручное подключение к  
охлаждению**

«Шнуровка — наружу!»

**Маленькие**  
охлаждающие  
пластины и  
соединительные  
трубки между ними

**Гидравлический  
бэкплэйн**

**(автоматическое  
подключение к  
охлаждению)**



## Врожденные пороки всех **закрытых** систем жидкостного охлаждения

- ❑ **Жесткое решение:** ровно под одно «железо»
  - ☆ Защита инвестиций?
- ❑ **Сложность конструкции**
  - ☆ Надежность?
  - ☆ Цена?
- ❑ **Риски протечки** — в одном шкафу живут:
  - ☆ Вода и 512 гидравлических разъемов
  - ☆ Электричество (100 КВатт)
- ❑ **Проблема «точки росы»**
- ❑ **Неуправляемость и неоптимальность охлаждения**
  - ☆ Полное отсутствие адаптивности
  - ☆ Как тогда с PUE?...
- ❑ **Большой вес**

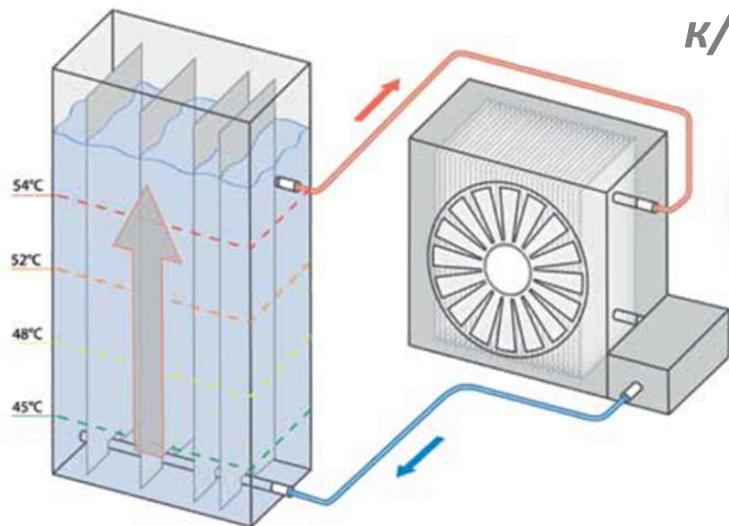




# Технология непосредственного жидкостного охлаждения IMMERS®



*«Все мы вышли из воды и должны туда же вернуться. В эту среду»  
к/ф «Национальные особенности рыбалки»*





# IMMERS: избавляем от проблем **закрытых** систем жидкостного охлаждения

- ❑ **Жесткое решение:** ровно под одно железо

- ☆ Защита инвестиций?

- ❑ **Сложность конструкции**

- ☆ Надежность?

- ☆ Цена?



- ❑ **Риски протечки** — в одном шкафу живут:

- ☆ Вода и 512 гидравлических разъемов

- ☆ Электричество (100 КВатт)

- ❑ **Проблема «точки росы»**

- ❑ **Неуправляемость и неоптимальность охлаждения**

- ☆ Полное отсутствие адаптивности

- ☆ Как тогда с PUE?...

- ❑ **Большой вес**

- ❑ **Построил инфраструктуру?**

Погрузить в нее можно что угодно!

- ❑ **Проще не бывает!**

- ☆ Надежность!

- ☆ Цена!



- ❑ **Нет рисков протечки! Все уже протекло!**

- ❑ **Нет «точки росы», нет требований к климату вокруг стойки**

- ❑ **Врожденная адаптивность**

- ❑ **Меньший вес**



# Охлаждающие жидкости

Жидкость	Характеристики	Токсичность
3M Fluorinert	1855 кг/м <sup>3</sup> 2040 Дж/(л К) 0,065 Вт/(м К)	Попадание в дыхательные пути, в глаза и на кожу
Green Revolution ElectroSafe	850 кг/м <sup>3</sup> 1190 Дж/(л К) 0,140 Вт/(м К)	Не токсична
Hardcore Computers CoreCoolant	770 кг/м <sup>3</sup> 1655 Дж/(л К) 0,137 Вт/(м К)	Попадание в глаза и на кожу
СТОРУС- ОХЛ14	911 кг/м <sup>3</sup> 1462 Дж/(л К) 0,124 Вт/(м К)	Не токсична



Охлаждающая жидкость СТОРУС-ОХЛ14 имеет российские ТУ, гигиенический и пожарный сертификаты.



## Нетоксичные охлаждающие жидкости

### Теплоемкость (на литр)

Green Revolution ElectroSafe  
СТОРУС-ОХЛ14

81.4%

100.0%

### Теплопередача

Green Revolution ElectroSafe  
СТОРУС-ОХЛ14

100.0%

88.6%

### Общая эффективность

Green Revolution ElectroSafe  
СТОРУС-ОХЛ14

91.9%

100.0%

Реальное достижение **PUE = 1.037**



# **2011–2014: Отечественный интерконнект, как осознанная необходимость...**





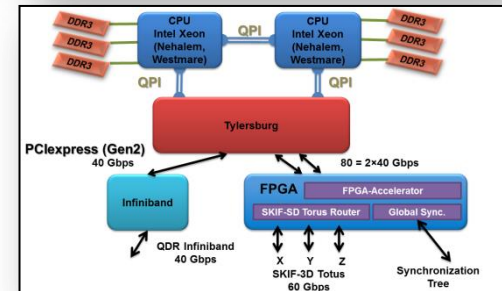
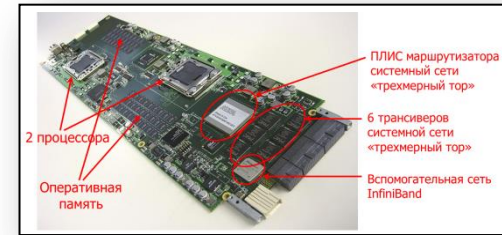
# 2009–2010: интерконнект SKIF-3D-Torus

## Перестраиваемая системная сеть

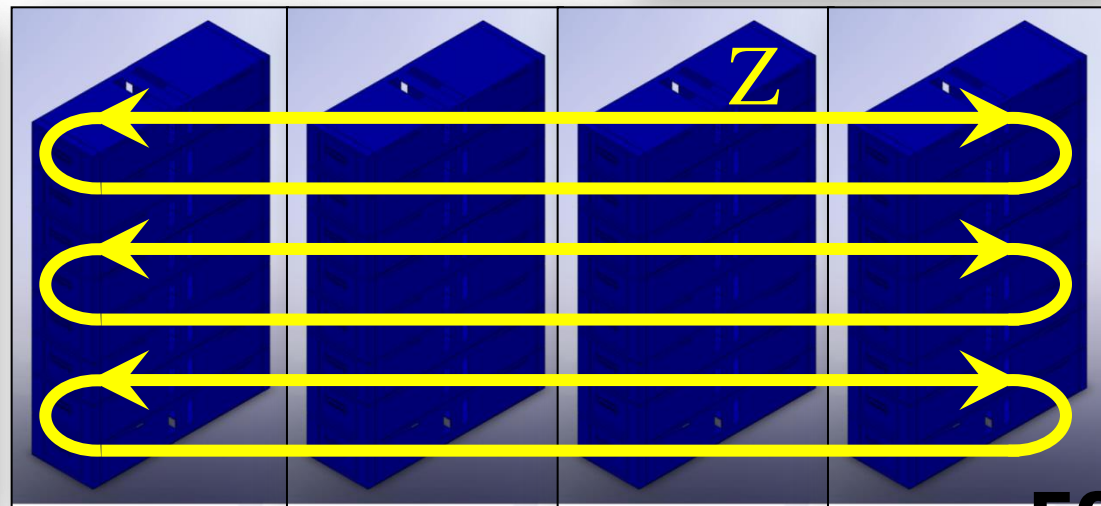
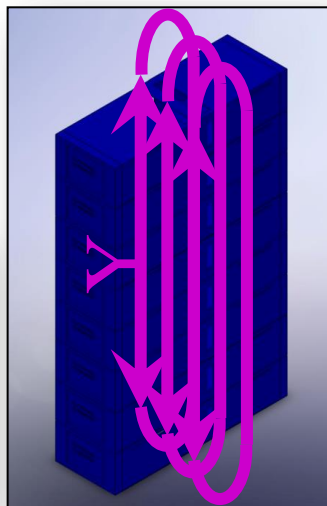
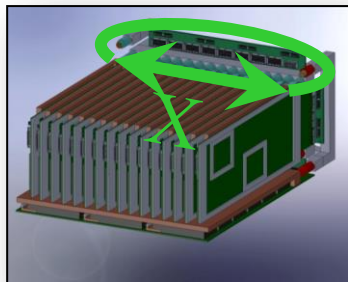
- Коммутаторы для конфигурирования сети
- Тор можно разбить на подтопы
- Задаче можно назначить свой собственный подтор

## Роутер реализован в FPGA. Поддерживает

- топологию 3D-тора
- отсутствие проблем вида deadlocks, livelocks, starvation
- виртуальные каналы и туннели
- Fault-tolerance, Multicasting, Broadcasting



**EUROTECH**  
Imagine. Build. Succeed.







# Архитектура стека ПО SKIF-3D-Torus

Прикладная программа

Коммуникационные библиотеки

SKIF-MPI

SKIF-SHMEM

SKIF-ARMCI

SKIF-GASNET

ALT Linux SKIF Cluster

SKIF-Driver

Коммуникационная библиотека  
нижнего уровня

SkifCh

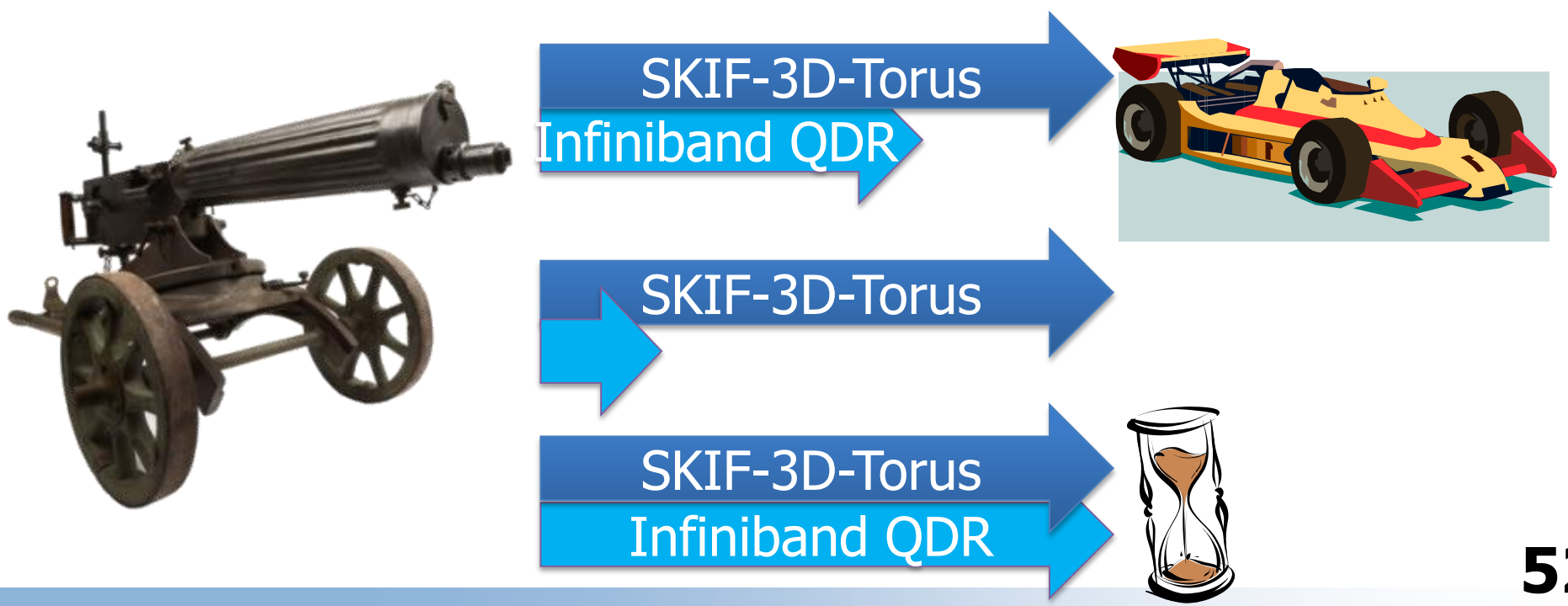
Маршрутизатор системной коммуникационной сети

SKIF-3D-router (VHDL)



# SKIF-3D-Torus vs. Infiniband QDR

	SKIF-3D-Torus	Infiniband QDR
Bandwidth (Gbps per node)	60	40
Message Rate (MT/s)	14	3
Latency ( $\mu$ s)	1–1.5	1–1.5



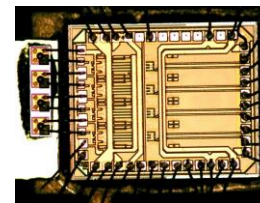
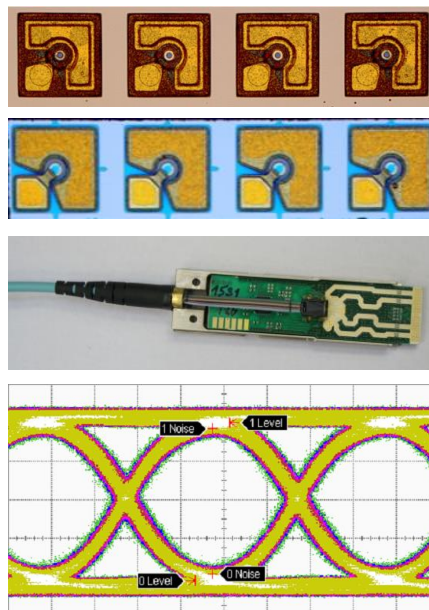
## 2013–2014: интерконнект «Паутина»



- Connector Optics Co.**  
**QSFP+ Активный Оптический Кабель**  
 56 Gbps = 4×14Gbps, array of vertical cavity lasers, array of PIN photodiodes, Chip, PCB, QSFP+ Connector, Cable

- ИПС РАН**  
**сетевой адаптер «Паутина»**  
 PCI-Express Gen3 ×8,  
 FPGA Altera Stratix V,  
 4×QSFP+,  
 224 Gbps = 4×56 Gbps

- ИПС РАН**  
**Pautina Software Stack**





# Архитектура стека ПО «Паутина»

**Прикладная программа**

**Коммуникационные библиотеки**

**SKIF-MPI**

**SKIF-SHMEM**

**...**

**Разрабатываемая  
библиотека Centaur**

**OS Linux**

**SKIF-Driver**

**Коммуникационная библиотека  
нижнего уровня**

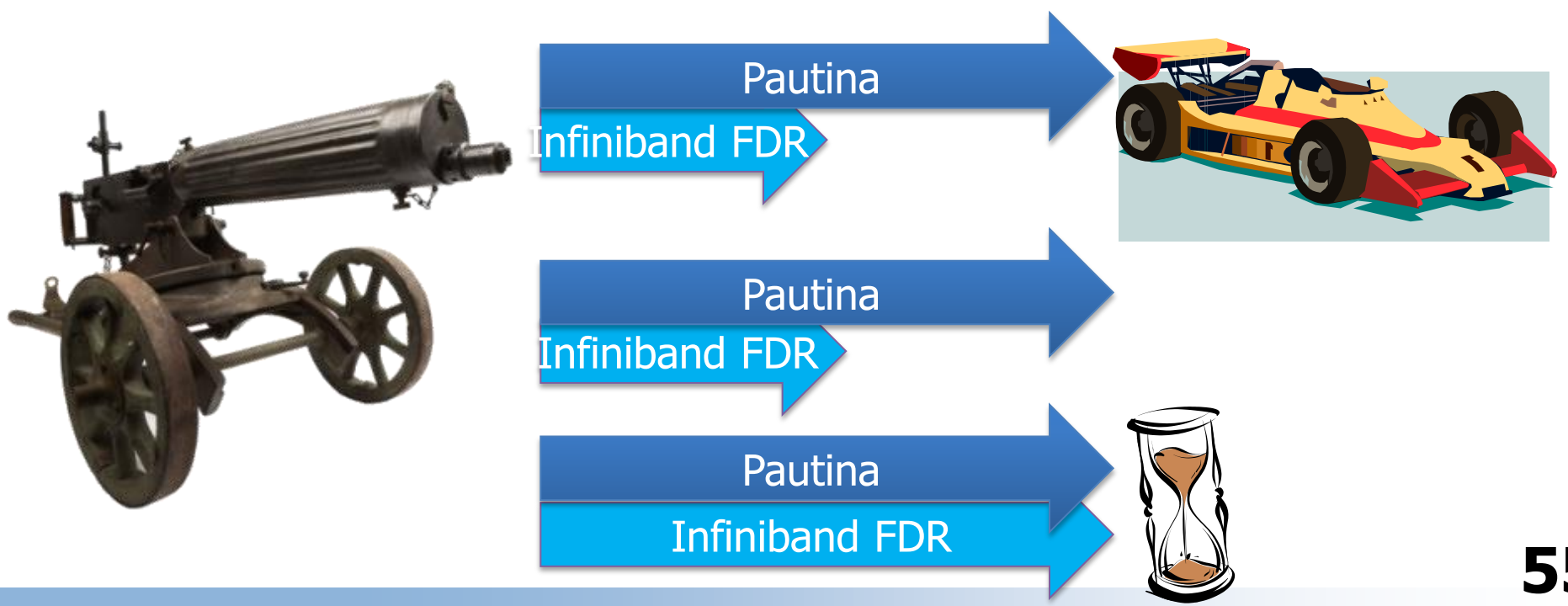
**SkifCh**

**Маршрутизатор системной коммуникационной сети**



# Паутина vs. Infiniband FDR

	Паутина	Infiniband FDR
Bandwidth (Gbps per node)	$4 \times 56 = 224$	$2 \times 56 = 112$
Message Rate (MT/s)	75	30
Latency ( $\mu$ s)	1–1.3	1–1.3





# 2011–2014: Организационная активность...





## 2010–2014: Опыт, сын ошибок трудных...

### 2010–2011: Опыт организации и проведения МСКФ

- Не самостоятельное мероприятие
- Один день
- Соблазны Москвы
- Трудно работать в «чужом доме»

### 2010–2011: Опыт создания НСТП

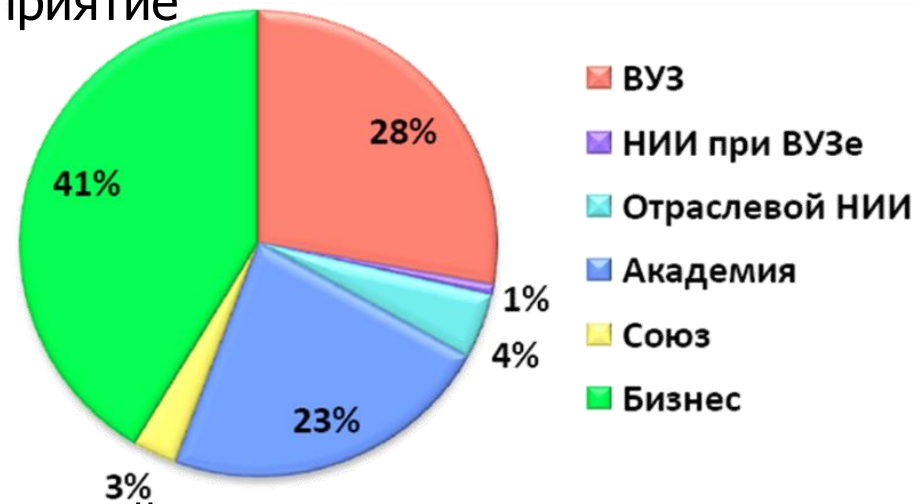
- Сегодня 260 организаций
- А в целом... очень поучительный опыт!

### 2012: Национальный суперкомпьютерный форум

- Важнейшее ежегодное общепромышленное мероприятие

### 2014: АНО «НСКФ»

- Постояннодействующая общепромышленная структура

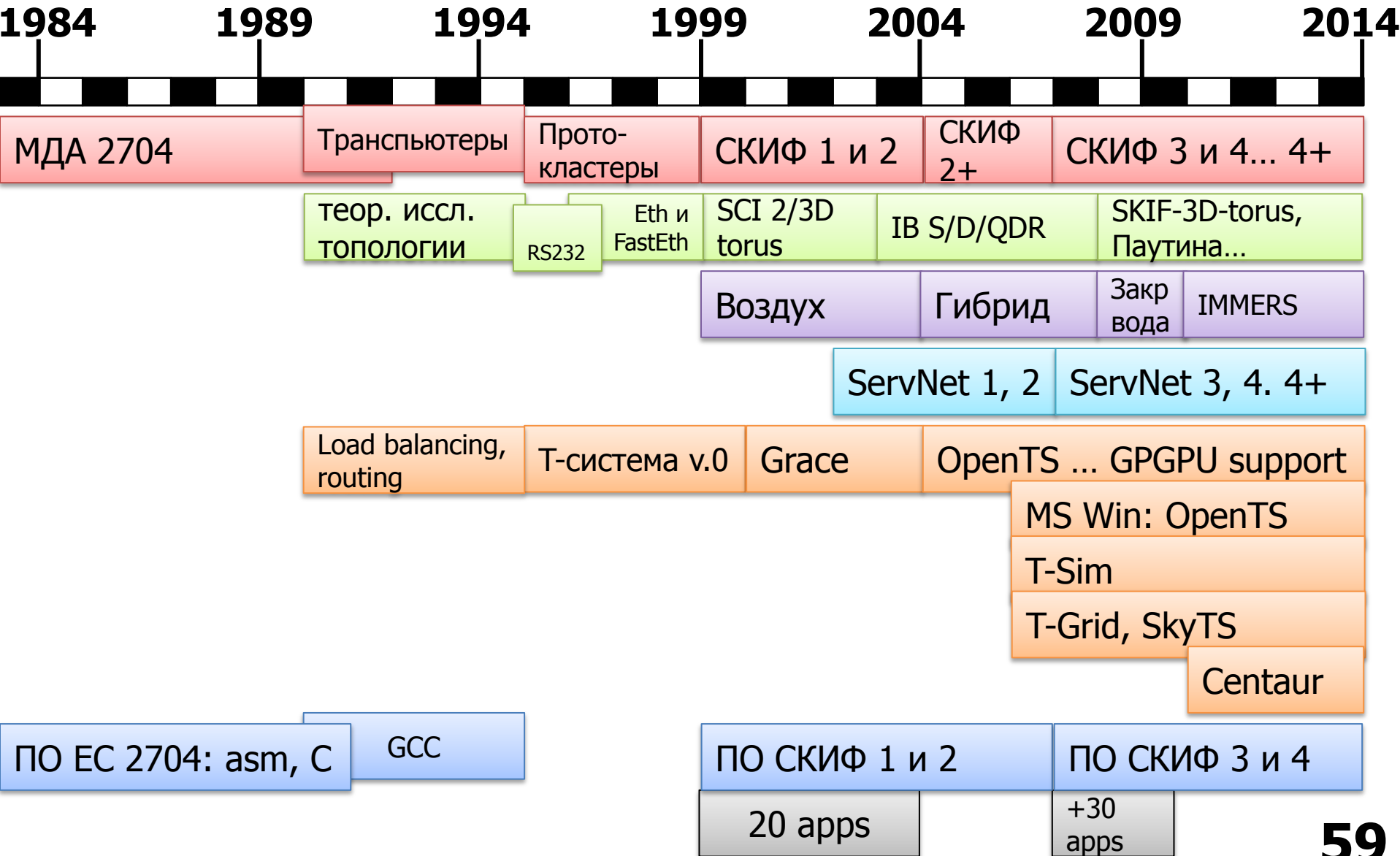




# Подводя итоги...









## Направления работ и научных интересов

### Архитектура

- Вычислительные узлы; стандартные CPU и ускорители: FPGA, GPU...
- Стандартный и перспективный интерконнект:
  - ☆ SKIF 3D-тор, «Паутина»
  - ☆ Синхронизация, коллективные операции и т.п.
- Перспективное охлаждение
- Сервисная сеть: SKIF ServNet v.1 ... v.4
- Цифровые вторичные блоки питания
- Системы хранения данных, активные хранилища

### Системное ПО

- ОС AltLinux SKIF-cluster edition, ППО
- стек ПО SKIF-MPI, SKIF-MON, СПОКУ, OpenTS, T-Sim, Centaur: гибридные «шаблонные» вычисления

### Инструментальное и прикладное ПО

- ИНС, обработка изображений, машинное зрение
- ИИ-обработка текстов
- прикладные задачи (восстановление испорченных изображений)

### Научно-организационная работа:

- МСКФ-2010, МСКФ-2011
- НСТП
- НСКФ-2012, НСКФ-2013, НСКФ-2014...
- АНО «НСКФ»