НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«КУРЧАТОВСКИЙ ИНСТИТУТ»

**Отделение
суперкомпьютерных систем
и параллельных вычислений**



Исследование производительности и применимости отечественной высокоскоростной коммуникационной сети Альфа при решении задач на суперэвм

Современные решения в области интерконнекта для HPC

InfiniBand (IB)

Omni-Path

HPE Slingshot

Ethernet для HPC (в т.ч. RoCE)

Top / mesh-сети (пример: Fujitsu Tofu)

ВКС «Альфа»

СМПО

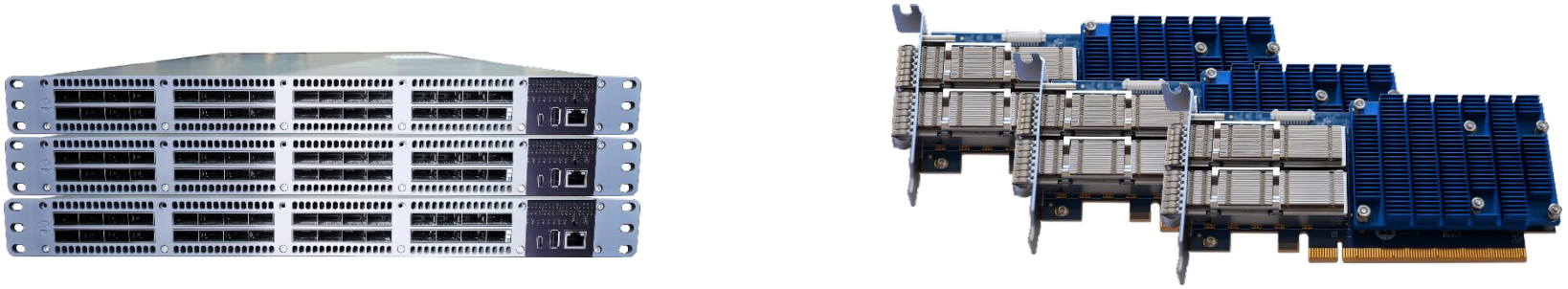
Зарубежные сети:

- До **800 Гбит/с** для InfiniBand XDR
- До **400 Гбит/с** для Slingshot 400(**200 Гбит/с** для Slingshot-11 (2-е место в TOP-500)
- До **400 Гбит/с** Cornelis Networks Omni-Path

Отечественные сети:

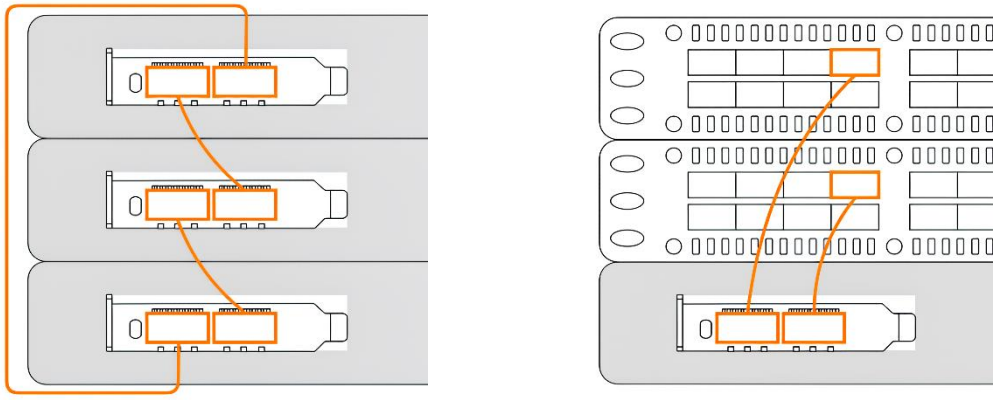
- СМПО – до **40 Гбит/с**
- ВКС Альфа – до **80 Гбит/с**

Аппаратные решения ВКС Альфа



(а) (б)

Рисунок 1 (а) – Коммутаторы и (б) – адаптеры ВКС Альфа



(а) (б)

Рисунок 2 (а) – Пример объединения адаптеров ВКС Альфа в кольцо из трёх узлов. (б) – Пример отказоустойчивого соединения адаптера и коммутаторов ВКС Альфа

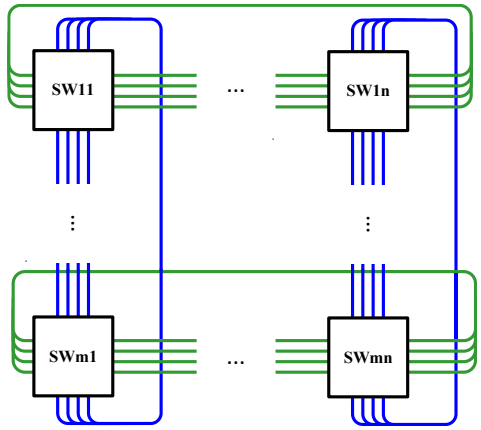


Рисунок 3 – Пример объединения 4-х коммутаторов ВКС Альфа в двухмерный тор

Методика исследования производительности

Основная задача

Исследование производительности коммутационной сети и определение влияния характеристик сетей (задержка, пропускная способность, масштабируемость, устойчивость к нагрузкам) на производительность приложений и микробенчмарков в условиях HPC.

Исследования производительности проводится в два этапа:

1. Синтетические тесты:

– MPI Uniband и MPI PingPong использовались для измерения латентности и пропускной способности;

– Linpack Benchmark использовался для оценки пиковой и устойчивой производительности вычислительной системы с учётом влияния сетевого взаимодействия.

2. Работа ПО КМ «Лазурит», реализующего вычисления методом RANS/ILES на блочно-структурированной расчетной сетке в суперкомпьютерной среде.

Аппаратные решения для тестирования

Исследование проводилось с помощью синтетических бенчмарков и ПО КМ «Лазурит».
на двух высокопроизводительных вычислительных системах:

- Выделенный стенд высокоскоростной коммуникационной сети «Альфа» (Далее – Стенд Альфа);
- Суперкомпьютер МВС-10П, установленный в НИЦ «Курчатовский институт» (Далее – Стенд НИЦ КИ).

Таблица 1. Характеристики вычислительных узлов

Наименование стенда	Стенд Альфа	Стенд НИЦ КИ
Процессор и/или сопроцессор	Intel Xeon Gold 6336Y	Intel Xeon Platinum 8358
Базовая частота процессора	2,4 ГГц	2,6 ГГц
Турбо-частота	3,6 ГГц	3,4 ГГц
Количество физических ядер в процессоре	24	32
Количество виртуальных ядер в процессоре	48	64
Теоретическая пиковая производительность процессора, ТФлопс	1,84	2,66
Количество процессоров в узле	2	2
Количество виртуальных ядер процессоров в узле	96	128
Теоретическая пиковая производительность в узле, Тфлопс	3,68	5,32
Кэш процессора 3-го уровня, Мбайт	36	48
Тип и частота оперативной памяти	DDR4,1600 МГц (3200MT/s)	DDR4, 1600 МГц (3200MT/s)
Объем оперативной памяти в узле	512 Гбайт	256 Гбайт
Сеть межузлового обмена (интерконнект)	ВКС «Альфа»,80 Гбит/сек.	Omni-Path, 100 Гбит/сек.
Тип сетевой файловой системы СХД	NFS v.4	NFS v.4
Сеть подключения к СХД, Гбит/сек	Ethernet, 25	Omni-Path, 100

Результаты синтетических тестов

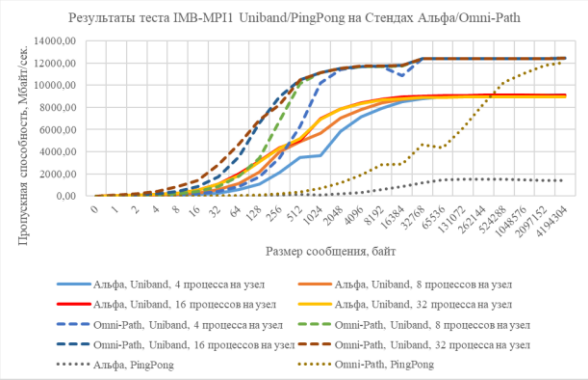


Рисунок 6 (а) – Пропускная способность сети IMB-MPI1 Uniband / PingPong на Стенде Альфа и Стенде Omni-Path

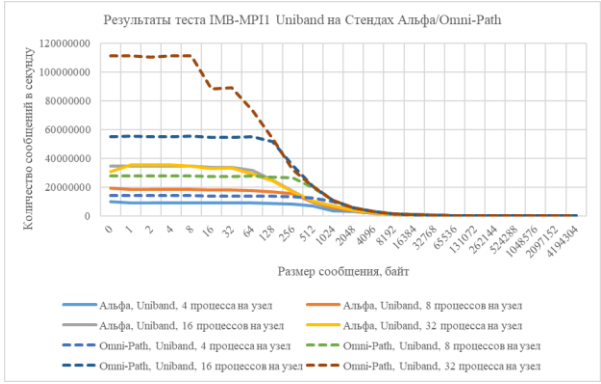


Рисунок 6 (б) – Темп выдачи сообщений в тесте IMB-MPI1 Uniband на Стенде Альфа и Стенде Omni-Path

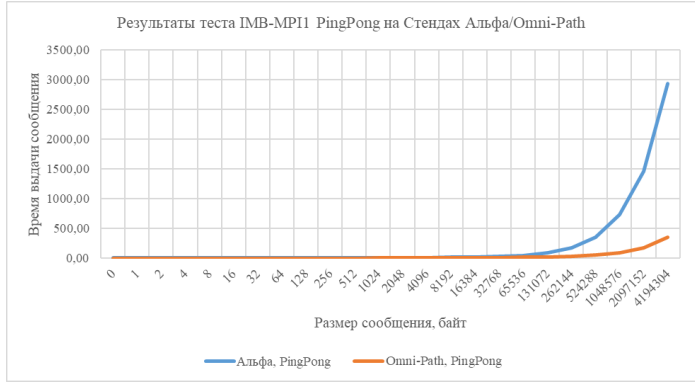


Рисунок 7 – Длительность теста IMB-MPI1 PingPong на Стенде Альфа и Стенде Omni-Path

Результаты тестов HPL

Кол-во узлов	1	2	4	5
Стенд Альфа, Rmax				
Производительность, ТФ	2,59	4,81	9,71	12,16
% от Rpeak	70,29	65,40	65,97	66,09
Масштабирование, крат	1,00	1,86	3,75	4,70
Стенд Альфа, 1,8 ГГц, Rmax				
Производительность, ТФ	2,49	4,94	9,79	12,18
% от Rpeak (2,7 ТФЛОПС)	90,21	89,40	88,55	88,10
Масштабирование, крат	1,00	1,98	3,93	4,88
Стенд Omni-Path, Rmax				
Производительность, ТФ	3,18	6,23	12,64	15,78
% от Rpeak	86,40	84,65	85,85	85,75
Масштабирование, крат	1,00	1,96	3,97	4,96

Таблица 2. Масштабирование теста HPL

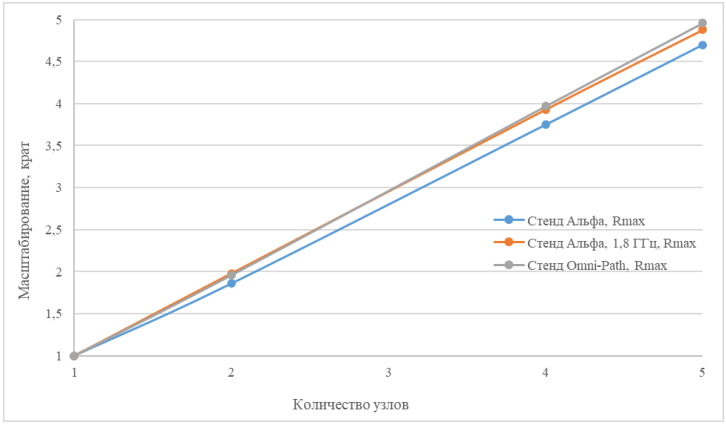


Рисунок 8 – Масштабирование теста HPL

Результаты тестирования ПО КМ Лазурит на Стенде Альфа и Стенде Omni-Path

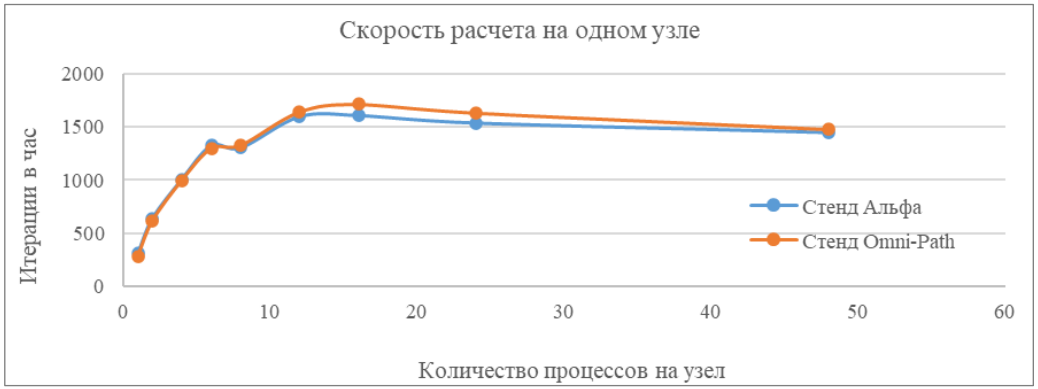


Рисунок 10 – Скорость расчета на 1 узле в зависимости от количества процессов

Таблица 4. Ускорение при расчёте на нескольких узлах, сводные показатели

Вычислительный кластер и количество процессов	2 узла	4 узла	5 узлов
Стенд Альфа, 12 процессов	1,93	3,92	4,86
Стенд Omni-Path, 12 процессов	2,05	3,92	4,87
Стенд Альфа, 16 процессов	1,93	3,68	4,63
Стенд Omni-Path, 16 процессов	2,01	3,81	4,80
Стенд Альфа, 48 процессов	1,94	3,69	4,17
Стенд Omni-Path, 48 процессов	1,92	3,62	4,38

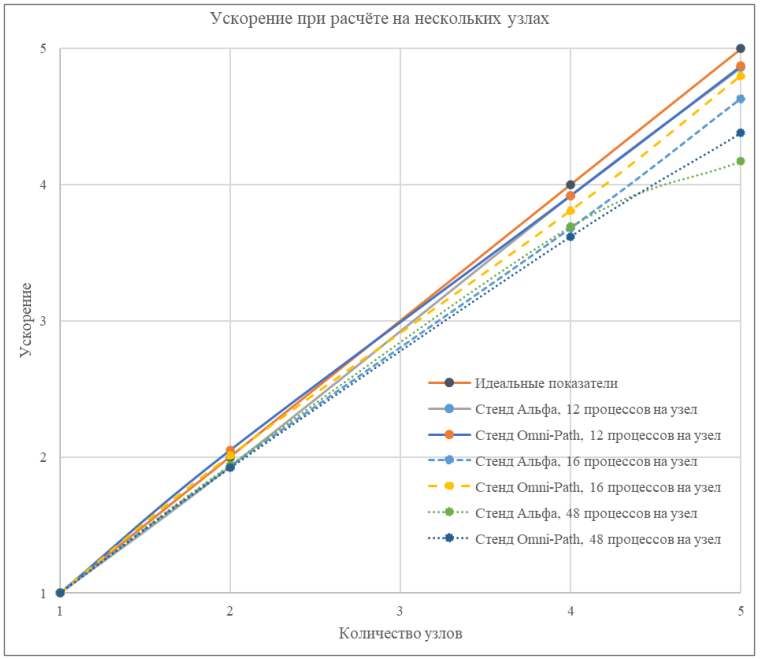


Рисунок 11 – Прирост скорости при расчёте на нескольких узлах, 12, 16, 48 процессов на узел

НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«КУРЧАТОВСКИЙ ИНСТИТУТ»

Отделение суперкомпьютерных систем и параллельных вычислений



Спасибо за внимание!