Outline

• Introduction
• Past and Ongoing projects
• RSC Tornado solution
• RSC PetaStream Performance
Development of innovative ultra-high density energy efficient HPC solutions delivering unique features and addressing specific end-user needs

Cutting-edge supercomputers and data centres for demanding customers
Strong Market Position

• **Leading** HPC solution provider in **Russia/CIS**

• Key installations in Russia according to **local Top50 list** - **5 systems** with RSC liquid cooling, **2.5+ PFLOPS** in total

• **One of the key** HPC players in **EMEA**

• In **Top10 worldwide** by Top500 list (#9)

• **First two Intel® Xeon Phi™ projects** in Europe and outside the USA (ranked by Top500)

Source: Top500.org (November, 2014)
• RSC’s share at **Top500** - **44%** of all **Russian systems** (incl. #81, Saint Petersburg Polytechnic University)

• RSC’s share at **Green500** - **44%** of all **Russian systems** (incl. #21, Saint Petersburg Polytechnic University)

• **The most energy efficient system in Russia** according to Green500 (Saint Petersburg Polytechnic University)

• **Leading Russian HPC player in Green500** - **Top4** the most energy efficient Russian systems out of 9 there
• Leading **computing density** per rack (PetaStream) – **1.2 PFLOPS**

• Leading **performance density** (PetaStream) – **560+ TFLOPS/m^3**

• Leading **power density** per rack (PetaStream) – **400+ kW**

• Highest **computing density** per rack* (Tornado) – **269.4 TFLOPS**

• Highest **performance density** (Tornado) – **210 TFLOPS/m^3**

• Highest **power density** per rack* (Tornado) – **100 kW**

• Leading **PUE = 1.057** (measured at customer site)

* 80x80cm 42U rack
Top projects

- **Saint-Petersburg State Polytechnic University**, 1.1 PFLOPS
- **Russian Academy of Sciences** (JSCC RAS), 600 TFLOPS
- **South-Ural State University** (SUSU), 473.6 TFLOPS
- **Russian Weather Forecast Agency** (Roshhydromet), 35 TFLOPS
- **Moscow Institute of Physics and Technology** (MIPT), 83.5 TFLOPS
- **Aviation Industry, Energy sector, Computer Graphics, Oil&Gas**
- … and many others

**over 2.5 PFLOPS**
of installed base of innovative liquid cooled supercomputers
The first Intel® Xeon® E5-2600 v3 based cluster in Russia and CIS
Peak performance over 1.1 PFLOPS
The most advanced technology and one of top powerful
RSC’s direct liquid cooling for energy efficiency at scale

Solution components

**RSC Tornado**
- Performance - 830 TFLOPS
- Recent Intel Xeon E5-2697 v3 CPUs
- Intel® S2600KP Server boards
- Intel® SSD DC S3500/P3700
- RAM DDR4 – 64-128 GB per node
- VDI & I/O Expansion packs

**RSC PetaStream:**
- Performance - 270 TFLOPS
- Intel® Xeon Phi™ 5120D
- Service processors Intel Xeon E5-2600
- Intel® SSD DC S3700
- FDR Infiniband
- **Emerson 400V DC** energy efficient power supply

- Shared fat tree Infiniband FDR (56 Gb/s) FBB fabric
- Shared parallel storage - 1 PB (Lustre) + 0,5 PB cloud storage
JSCC RAS upgrade (2014)

- The first project on Intel® Xeon Phi™ 7120D
- Aggregate peak performance ~600 TFLOPS
- One of top Russian HPC centers
- RSC’s direct liquid cooling technology

Solution components

**MVS-10P supercomputer:**
- 208 nodes ~ 2.5 TFLOPS each
- Dual CPU Xeon E5-2690
- Dual co-processor Intel Xeon Phi SE10X
- 64 GB RAM DDR3-1600 LV per node
- FDR Infiniband, fat-tree. NetApp & GPFS

**RSC PetaStream:**
- (co-)Processors - Intel® Xeon Phi™ 7120D
- Service processor Intel® Xeon® E5-2600
- Intel® SSD DC S3500
- FDR Infiniband
- Emerson 400V DC power supply
• First deployed in 2009, upgraded in 2011-2013
  384 Nodes
  – 2x Intel Xeon X5680 @ 3.33GHz (130W TDP)
  – 1x Intel Xeon Phi coprocessor SE10X (300W)
  – 24/48GB DDR3R-1333 RAM
  – QDR Infiniband, Fat-tree. Lustre FS
  – Panasas (including native mount on MIC)

• Ratings
  – #127 in Top500 (Nov’13)
  – #64 in Green500 (Nov’13), max #40
  – #2 in Russia by energy efficiency level
  – #4 in Top50 (Russia/CIS)

• Sizing
  – 7 compute racks, 3 support racks
  – 50 sq. m./350KW
  – 473.6 TFLOPS Rpeak,
  – 288.2 TFLOPS Rmax (HPL)
  – 980 MFLOPS/W
"Both SUSU and JSCC RAS are state of the art high performance computing centers with competent staff running the highly ranked Top500 and Green500 powerful and energy efficient supercomputers. The facilities both use RSC Tornado based systems with innovative liquid cooling and newest Intel Xeon Phi coprocessors which provide impressive high performance capabilities and energy efficient solutions to solve very demanding science research and engineering problems." – said Jack Dongarra, the world famous HPC expert.
Other names and brands may be claimed as the property of others.

**RSC Technology Evolution**

- **2009-10**
  - 35 TFLOPS
  - 52 kW per rack
  - 96 nodes

- **2011-today**
  - 66.3 TFLOPS
  - Up to 70 kW per rack
  - 72 heterogeneous nodes

- **2012-today**
  - 211 TFLOPS
  - Up to 100 kW per rack
  - 128 homogeneous nodes

- **2013-today**
  - 1.2 PFLOPS
  - 400+ kW per rack
  - 1024 nodes

**World Record**
Two worlds. State of Art.

RSC TORNADO CLUSTER SOLUTION

RSC PETASTREAM MASSIVELY PARALLEL SUPERCOMPUTER
RSC Tornado cluster solution

**Processor**
- Intel® Xeon® E5-2600 v3
- Up to 18 cores x86 with 36 threads
- 0.66 TFLOPS (Peak)
- 68 GB/s peak mem. BW
- 45 MB shared cache

**Compute node**
- Two compute chips
- Up to 128GB of DDR4-2133 RAM
- 1 integrated IB (FDR/QDR) up to 56 Gb/s
- 1 addon PCIe Gen3 x16 IO card
- Direct liquid cooling of all components
- OS: Linux, Windows
- Virtualization: Hyper-V, VMware, Xen, KVM

**RSC Tornado Expansion Pack:**
- HPC Expansion Pack
- BigData Expansion Pack
- VDI Expansion Pack
- Security & Protection Expansion Pack
- Connectivity Expansion Pack
- SDM Expansion Pack
- and more, available by request

**Complete solution**
- Scales to many PFLOPS
- Modular: tailored to customer’s needs
- Flexible network options
- Based on COTS components

**Compute cabinet**
- Up to 269 TFLOPS Peak performance or 128 fastest servers
- Intel Xeon E5-2600 v3 family processors, v4 family upgrade ready
- Fully integrated software stack for HPC and Cloud “RSC BasIS”
- Single System Management and Monitoring Point
- 0.64m² / 6.9 ft² footprint

*Other names and brands may be claimed as the property of others.*

Copyright © 2009, Intel Corporation.
RSC Tornado Extensions

- **HPC Expansion Pack**: up to 2 Co-processors and IB
- **VDI Expansion Pack**: fast PCIe SSD and 3D video
- **BigData Expansion Pack**: enhanced storage, performance/$
- **Security&Protection Expansion Pack**
- And others...
RSC PetaStream

1.2 PFLOPS per rack

Massively-parallel ultra-high density*

HPC solution

* Performance density over 560 TFLOPS/m³
RSC PetaStream outline

**Processor**
- Intel® Xeon Phi™
- 61 x86 cores / 244 threads
- > 1.2TFLOPS peak perf.
- 352 GB/s peak mem. BW
- 30 MB shared cache

**Compute node**
- Intel® Xeon Phi™ 7120D
- One compute chip
- 16GB of RAM
- 64Gbps IO bandwidth
- Linux µOS

**Compute module**
- 8 compute nodes
- Over 200 Gbps external IO bandwidth
- Direct liquid cooling of all components
- Integrated node management
- Effective DC 400V power system

**System**
- Path to ExaScale
- Proven RSC Direct Liquid Cooling Technology
- Scalable/modular: tailored to customer’s needs
- Flexible network options
- Based on COTS components

**Cabinet**
- Over 1.2PFLOPS peak performance
- 250K threads / 1024 nodes
- Up to 400 kW
- Integrated management
- 1m² / 10.8 ft² space
RSC PetaStream™ Cabinet

• Scalable modular high-density system design
  ➔ Rack dimensions: 1x1x2.2m / 48U => 2.2 m³, 1 m²
  ➔ 1024 RSC PetaStream compute nodes are grouped into 128 modules with 0.5PB local storage
  ➔ Integrated GigE switches for management networks
  ➔ Integrated node management
  ➔ Efficient DC power distribution within a cabinet
    ▪ up to 450kW power consumption (peak)
    ▪ Emerson 400V DC power supply system
  ➔ Integrated coolant distribution
    ▪ low pressure, low requirements to liquid quality
  ➔ Optimized weight

• External aux equipment (modular)
  ➔ AC/DC power converters with (optional) UPS
  ➔ (optional) liquid-to-liquid heat exchanger with industrial control system for internal loop
  ➔ Liquid-cooled rack for low-power air-cooled equipment, such as Infiniband network switches, storage, etc.

1.2 PF peak with ~250K threads in 1 rack
Other names and brands may be claimed as the property of others.

Copyright © 2009, Intel Corporation.

RSC PetaStream delivers high computational density and energy efficiency with record-level performance for ExaScale.

and $\frac{1}{2}$ energy to solution

$= 9x$
<table>
<thead>
<tr>
<th>Code/app</th>
<th>Developer/partner</th>
<th>Area/Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAGMA MIC</td>
<td>ICL University of Tennessee</td>
<td>Dense linear algebra</td>
</tr>
<tr>
<td>AstroPhi</td>
<td>Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia</td>
<td>Astrophysics, gas dynamics</td>
</tr>
<tr>
<td>GROMACS</td>
<td>St-Petersburg Polytechnic</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>CACTUS</td>
<td>University of Lousiana</td>
<td>Astrophysics</td>
</tr>
<tr>
<td>GAMESS</td>
<td>Chem. Dept. Lomonosov MSU</td>
<td>Quantum Chemistry</td>
</tr>
<tr>
<td>Pikador</td>
<td>NNSU</td>
<td>Patrice-In-Cell</td>
</tr>
</tbody>
</table>
Communicating agents

[Diagram of a network system with CPUs, memory, PClie switches, and Infiniband connections. The diagram shows data flow and network topology.]
MPI Point-to-Point bandwidth

MPI Bandwidth Tests, OSU v4.2

- PetaStream[near]
- PetaStream[far]
- JSCC[near]
- JSCC[hosts]
- JSCC[far]
MPI latency – small (1B-512B) messages
MPI Message rate

MPI Message Rate Test, OSU v4.2

Number of rank pairs

Messages/second

PetaStream[near]
PetaStream[far]
JSCC[near]
JSCC[hosts]
JSCC[far]
MAGMA MIC Performance

Performance (GFLOPS)

Matrix size $N=M$

- 2x Intel Xeon E5-2690
- PetaStream, 1 Intel Xeon Phi 5120D
- PetaStream, 2 Intel Xeon Phi 5120D
- PetaStream, 3 Intel Xeon Phi 5120D
- PetaStream, 4 Intel Xeon Phi 5120D
- PetaStream, 5 Intel Xeon Phi 5120D
- PetaStream, 6 Intel Xeon Phi 5120D
- PetaStream, 7 Intel Xeon Phi 5120D
- PetaStream, 8 Intel Xeon Phi 5120D

* Other names and brands may be claimed as the property of others.

Copyright © 2009, Intel Corporation.
Other names and brands may be claimed as the property of others.
GROMACS performance testing

Alexey Shvetsov, FSBI Petersburg Nuclear Physics Institute, NRC Kurchatov Institute, GROMACS Developer
Evgeniy Petukhov, St.Petersburg State Polytechnic University
Problem description

Nucleosome and its different assembly states (such as hexasome, tetrasome) in water (approx. 1M atoms). Nucleosomes are essential structural and regulatory element for DNA packaging in eukaryotic cell nuclei. Such simulations help better understanding of fundamental mechanisms of gene transcription regulation.

System features: large amount of volume distributed point charges (~1/6rd of total number of atoms)
Nucleosome

Nucleosome

Tetrasome
GROMACS native version compilation on Xeon Phi

Version: 5.1-dev (git master branch)
Patches: bitmap (allows arbitrary number of omp threads)
Tools and libraries: Intel® Cluster Studio XE 2015 (C, C++, MKL, MPI)
Standard build options
Performance results

Being tested on Jaguar RF showed better scalability (up to tens of thousands of cores), but for this case PME gives more performance.
• Innovative implementation of massively parallel architecture using the best available industrial standard components

• World record of computing density of 1.2 PFLOPS peak performance per rack and 560 TFLOPS/m³

• 250,000 execution threads in a single rack

• Significant step ahead to create Exascale level supercomputers

• Possibility to develop and optimize applications for future massively parallel systems

• Software and hardware investment protection towards the Exascale era
Thanks

www.rscgroup.ru