SCHEME: Software Abstractions for Heterogeneous Multicore Systems

Per Stenström, Ulf Assarsson and Philippas Tsigas
Chalmers University of Technology

Divisions of Computer Engineering and Networks and Systems
Department of Computer Science & Engineering
High-Productivity Software Design in the Multi/Many-core Era

Plug & play

Productivity programming languages (e.g. C/C++, Java)

Focal Point of SCHEME
SCHEME: High-Productivity Software Stack for Many-cores

- Efficiency/Domain programmers
- Efficiency-only programmers
- Computer architects

Architecture-Aware & Domain-Specific Algorithms & Data Structures

Runtime with Parallelism Capabilities

Hardware Primitives (e.g. TM)

Increased level of abstraction
Transactional Memory: Unlocking Concurrency with low Programming Efforts

Traditional locking primitives: No concurrency and poor composability

```java
void transfer(A, B, amount) {
    synchronized(bank) {
        withdraw(A, amount);
        deposit(B, amount);
    }
}
```

Transactional memory: Can unlock concurrency, but poor perf. predictability

```java
void transfer(A, B, amount) {
    atomic {
        withdraw(A, amount);
        deposit(B, amount);
    }
}
```

Objective: Methods to improve programmability and perf. predictability
Major Achievements

- Innovative transactional memory concepts
  - Low programming effort
  - High performance predictability
  - Low complexity
- Real-time aware transactional memory concept
- Blaze memory compression: up to 9X more memory capacity

Significance:
- Intel released its first transactional memory enhanced product in June 2013 (Intel Haswell)!!
Lock-free Data Structures

- Fundamental building blocks of efficient and fault tolerant concurrent programs.

- Lock-free Binary Search Trees (LFBST): Efficient concurrent Binary Search Trees that provides progress guarantee even if some threads fail. (submitted to ICDCS 2015)
  - Light weight design that ensures much improved parallelism compared to the existing designs.
  - Theoretically improved complexity of $O(H + c)$ compared to $O(cH)$ of the existing LFBSTs – $H$ = height, $c$ number of contending threads.
Our design outperforms all the existing designs of LF BSTs.

Efficiently scales with increasing number of threads.

100% better performance than the state of the art ConcurrentSkipListMap available in Java concurrency library.
Computer Graphics Algorithms

- Fractal compression of highly detailed/complex geometry
  - ~2 orders of magnitude compression
  - No decompression needed
  - Computations as fast as or faster than on the uncompressed data.

- Used for:
  - real-time rendering (games)
  - Offline rendering (movies)
  - 3D scanned environments
**Fractal compression also in time**

- Enables: **Free Viewpoint Video** - even over Internet (req: 5-40Mbit/s)

1. Collect camera shots of the same scene, from ~10 different directions.

2. Number crunch input streams.
   Store free-viewpoint video in our new video format, for later playback from any position and angle.

3. Real-time playback from any view.

A real-time navigation example from Free-viewpoint video playback

Another real-time navigation example from same Free-viewpoint video
Exploitation - Collaborations and Startup

- **MindArk**
  - Hair rendering. Algorithm in product.

- **Fraunhofer FCC**
  - Collision detection. In product + joint paper

- **Bosch + Avalanche Studios**
  - 1K-1M lights. In products
  - Also joint conference courses/talks.
  - **Intel** - online graphics demos.

- **Activision “Call of Duty”**

- **Startup: Mindary**
  - Realistic Web 3D graphics.
  - Selling to EON Reality

Free Viewpoint Video.

**Simulation and Visualization of Hair for Real-Time Games**
**PDQ: Parallel Distance Queries for Deformable Meshes.**
**Many Many Light Sources in real time. 7 publications.**
**Fast, Memory-Efficient Construction of Voxelized Shadows.**
Exploitation - Blaze Memory

Memory compression technology partly funded by this project

Our unique position
- Offers up to 9X higher memory capacity
- Negligible performance loss
- Simple hardware, no software changes - “under the hood”

Exploitation plan
- Prototyping within the FP7 Euroserver project
- Business development based on Vinnova support through Innovationskontoret Väst
- Team: Si Valley based entrepreneur, one tech. lead and 4 technology developers

Value proposition
- Lower infrastructure cost
  - 20% is memory today
- Lower electricity cost
- Higher performance

Relevant for smartphones/tablets, laptops, servers, HPC systems

Compression factor for applications:

Ideal

Real
Career Development

Ph.D. and Lic Exams


- Two Ph.D.s are expected in 2016

Postdocs:

- Rubén Titos Gil, 2012-2014, Barcelona Supercomputing Center
- Erik Sintorn, 2014-, Chalmers
- Miquel Pericàs, 2014-, Chalmers
Key Numbers

Team
- 3 professors
- 3 post-docs
- 6 PhD students

Publications
- 25 publications
- 1 patent
- 1 spin-out (another one about to spin out)

Other project spin-offs
- 1 ERC advanced grant
- 1 Artemis project (w/ Volvo and Arccores)
- 2 European projects