In-Silico Science and Technology: From Atoms to Cognitive Computing

Dr. Alessandro Curioni
IBM Fellow
VP Europe and Director, IBM Research Zurich
IBM Research - Zurich

- Established in 1956
- 45+ different nationalities

- Open Collaboration:
  - FP7: 277 projects engaged, 68 funded, 1,900 partners
  - H2020: 52 applications, 341 partners

- Two Nobel Prizes
  (1986 and 1987)

- Binnig and Rohrer Nanotechnology Centre opened in 2011 (Public Private Partnership with ETH Zürich and EMPA)
Outline of the Presentation

• Introduction  
  *in silico* science and technology  
  the third pillar of innovation & Discovery

• Example  
  Battery 500  
  Development of Li-Air energy storage

• Example  
  SEMD  
  Towards billions of quantum (SCF) particle simulations

• Outlook  
  Dealing with Complexity: Cognitive
“Computational science is now indispensable to the solution of complex problems in every sector, from traditional science and engineering domains to such key areas as national security, public health, and economic innovation”

“Computational science has become the third pillar of the scientific enterprise, a peer alongside theory and physical experiment.”
HPC Evolution Enables Simulations of Realistic Models

ab initio quantum mechanics, many-body potentials, pair potentials

Simulation length scale: nanometers, micrometers
Simulation time: picoseconds, nanoseconds, microseconds, milliseconds

Today (0.1-10 petaFLOPs) vs. Tomorrow (exaFLOPs)
Algorithm Re-Engineering

Ab-initio Molecular Dynamics (CPMD)
Perfect scaling to 6 Millions of Threads (BG/Q)

Computational Fluid Dynamics
14.4 PFlops (72%) sustained performance on 6 Millions of Threads (BG/Q)

Using Millions of Threads on the BG/Q Supercomputer
Valery Weber, Costas Bekas, Teodoro Laino, and Alessandro Curioni
IPDPS 2014

14 PFLOP/s simulations of cloud cavitation collapse
Diego Rossinelli Babak Hejazialhosseini, Panagiotis Hadjidoukas, Costas Bekas, Alessandro Curioni and Petros Koumoutsakos,
SC13 - Gordon Bell Winner
Outline of the Presentation

• Introduction  
in silico science and technology  
the third pillar of innovation & Discovery

• Example  
Battery 500  
Development of Li-Air energy storage

• Example  
SEMD  
Towards billions of quantum (SCF) particle simulations

• Outlook  
Dealing with Complexity: Cognitive
Collaborative project aims to develop Lithium Air to boost the range of rechargeable batteries for all-electric cars from less than 100 miles (160 km) today to as far as 500 miles (804 km).
Method Development: highly scalable HFX

\[ E_x = -\frac{1}{2} \sum_{i,j=1}^{N} \int \int \psi_i^*(r_1) \psi_j(r_1) \frac{1}{r_{12}} \psi_i(r_2) \psi_j^*(r_2) d^3 r_1 d^3 r_2 \]

- Observe the double integral:
  - This means we need \(N^2\) FFT pairs (inverse/forward)
  - Thus the cost of HFX runs as \(O(N^2 M \log(M))\)

- FFTs become the main computational kernel
- Parallelizing to extreme scale is absolutely essential to tackle the excessive cost of HFX

- We focused on:
  - Efficient PW distribution schemes
  - Load balance of computations
  - Reduction of cost by means of thresholding: Do not calculate all integral pairs

BG/Q Sequoia @ LLNL. 96 Racks
1.6M cores, 6.4M threads
BAT500 – HPC-based Simulations and Computer Aided Design
Unveiling new pieces in the puzzle of Lithium-air-batteries

Activities and Results:

- electronic bandstructures of Li$_x$O$_y$ conductivity
- overvoltage – or lack thereof
- extensive study and forecasts of electrolyte solvent stability
- electron transport in Li$_2$O$_2$

Mostly done on a very large IBM Blue Gene at DOE Argonne National Lab

By the Computational Sciences team at IBM Research – Zurich

Thanks to INCITE awards

Simulations of Li$_2$O$_2$ in Propylenecarbonate, T. Laino, A. Curioni, A New Piece in the Puzzle of Lithium/Air Batteries, Chemistry, DOI 10.1002/chem.201103057 (22 February 2012)
Reactivity of Propylenecarbonate on Li$_2$O$_2$

On a Petaflop IBM Blue Gene at DOE Argonne Natl Lab (INCITE) ZRL and ARC systems

We extended this approach to screen different solvents with respect to their intrinsic stability versus Li$_2$O$_2$ solid particles.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Energy Barrier (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>– (*)</td>
</tr>
<tr>
<td>NMP</td>
<td>24</td>
</tr>
<tr>
<td>NMP-tBut</td>
<td>33</td>
</tr>
<tr>
<td>NMP-F3</td>
<td>– (*)</td>
</tr>
<tr>
<td>NMP-F6</td>
<td>– (*)</td>
</tr>
<tr>
<td>2Met-NMP</td>
<td>51</td>
</tr>
<tr>
<td>PEG-5</td>
<td>16</td>
</tr>
<tr>
<td>Met-PEG-5</td>
<td>23</td>
</tr>
<tr>
<td>F-PEG-5</td>
<td>40</td>
</tr>
<tr>
<td>CH3CN</td>
<td>35</td>
</tr>
</tbody>
</table>

A Li-ion conducting Membrane will be critical.

- Should be mechanically robust
- Resist Dendrite Penetration
- Be stable in Li-Air environment

Novel in-silico materials design

Summary (BATT500)

- Demonstrated solvent degradation to be the limiting factor for the building of Li/Air batteries

- Promising results from novel Li-ion solid state conducting electrolytes developing new materials

- Insight crucial for the development of really reversible Na/Air systems (1000 Cycles)

- Need of more effective scheme for in-silico screening of new materials

- Simulations crucial in steering the direction of the research effort
Outline of the Presentation

• Introduction  *in silico* science and technology
  the third pillar of innovation & Discovery

• Example  Battery 500
  Development of Li-Air energy storage

• Example  SEMD
  Towards billions of quantum (SCF) particle simulations

• Outlook  Dealing with Complexity: Cognitive
SEMD: Motivations

- Applications of quantum Hamiltonians to biological systems is limited by the cost of performing long calculations on large systems (> 30K atoms).
- Classical forcefields and QM/MM are good for conformational changes and localized reactions. Thus the need for developing scalable algorithms that allow the applications of quantum Hamiltonians to biological systems, to:

  - large scale ion motion
  - large scale electron transfer
Born-Oppenheimer molecular dynamics

Each MD step requires \( U \) to be calculated at the relaxed ground state electronic density.

\[
M_I \ddot{R}_I = -\frac{\partial U(R, \rho^{SC})}{\partial R_I},
\]

\[
\rho^{SC}(r, t) = \text{SCF}[\rho^{SC}(r, t - \Delta t)].
\]

Each SCF iteration requires the construction of the density matrix.

As a good trade-off of accuracy and computational speed, we employ a semi-empirical Hamiltonian (NDDO), but other types of Hamiltonian (HF, DFT, etc.) would be equally working.

The core operation of the SCF iterations is the sparse matrix-matrix multiplication.
What are the current limitations to go HUGE?

(DBCSR: Distributed Block Compressed Sparse Row)

Weak scaling for systems ranging from 32,928 atoms on 144 cores to 1,022,208 atoms on 5,184 cores. Communication and book-keeping have a contribution that grows with the square root of the number of MPI tasks (atoms).

The dashed line is a fit to the data using \( f(N) = a + b\sqrt{N} \).

(JCTC, 2012, 8, 3565−3573)
Semiempirical Molecular Dynamics (SEMD) I: Midpoint-Based Parallel Sparse Matrix–Matrix Multiplication Algorithm for Matrices with Decay

Valéry Weber,* Teodoro Laino,† Alexander Pozdneev,‡ Irina Fedulova,§ and Alessandro Curioni†

†IBM Research–Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland
§IBM Systems Lab Services Russia/CIS, Presnenskaya Nab., 10, 123317, Moscow, Russia
§IBM Science and Technology Center, Presnenskaya Nab., 10, 123317, Moscow, Russia

ABSTRACT: In this paper, we present a novel, highly efficient, and massively parallel implementation of the sparse matrix–matrix multiplication algorithm inspired by the midpoint method that is suitable for matrices with decay. Compared with the state of the art in sparse matrix–matrix multiplications, the new algorithm heavily exploits data locality, yielding better performance and scalability, approaching a perfect linear scaling up to a process box size equal to a characteristic length that is intrinsic to the matrices. Moreover, the method is able to scale linearly with system size reaching constant time with proportional resources, also regarding memory consumption. We demonstrate how the proposed method can be effectively used for the construction of the density matrix in electronic structure theory, such as Hartree–Fock, density functional theory, and semiempirical Hamiltonians. We present the details of the implementation together with a performance analysis up to 185 193 processes, employing a Hamiltonian matrix generated from a semiempirical NDDO scheme.

Weak Scaling

- Time per DM (density matrix) build wrt MPI tasks, PM6
- About 19 waters per task
- Parallel efficiency: 92% at 110592 MPI tasks (2.1M waters)
- Nbr non-zero elements: 1.6k/water (O1) and 1.0k/water (O2)

Constant walltime with proportional resources:
On largest BG/Q available we can simulate billions of particles at the quantum level…
Strong Scaling

- Time per DM (density matrix) build wrt MPI tasks, PM6
- 110k (S1), 373k (S2) and 1124k (S3) waters
- Speedup: 0.5 (S1), 1.7 (S2), and 3.7 (S3)

With OMP parallelization we can achieve a DM build in approx. 8 sec, on BG/Q (4 MPI / 8 OMP per node)
- Total communication volume (Isend/Irecv) per DM build wrt MPI tasks
- 110k (S1), 373k (S2) and 1124k (S3) waters
- BlueGene/Q
Our new development (MPSM3) shows:

- close to perfect weak scaling
- very good strong scaling
- communication volume decreases as nbr task increases

Ongoing work to ref-fit SEMD parametrization to provide high quality dynamic
Providing proportional resources, we can perform a MD step in less than one seconds regardless of system size.

Getting ready for pre-exascale systems (DCS)! Stay Tuned!
Outline of the Presentation

• Introduction  *in silico* science and technology
the third pillar of innovation & Discovery

• Example  Battery 500
Development of Li-Air energy storage

• Example  SEMD
Towards billions of quantum (SCF) particle simulations

• Outlook  Dealing with Complexity: Cognitive
What Next? Dealing with Complexity

Simulations

Input / Constraints

Big Data

Knowledge graph

Cognitive Computing

Output
(i.e. new material/device process conditions)
New frontier workloads: Cognitive Material Design

**Simulation**
- Theory + models + computation
- Structured & controlled
- Researcher bias

**Published knowledge**
- Much less bias
- Very rich
- Unstructured + noisy

**ARSENAL**
- Text analytics
  - $O(\text{day})$ per document page using 1 GPU
- Graph Analytics
  - $O(N^2)$ cost. Easily reaches Exascale per graph for large scale graphs needed
- Advanced machine learning & Statistics
  - $O(N^3)$ cost for Deep Neural Nets. Number of nodes scale with data size. Tens of millions of nodes needed
- Simulations
  - Even advanced methods (SEMD) need post Exascale for real life simulations
Extract Knowledge from Graph: Node Centrality and Sub-Graph Similarity

Translate Query to Subgraph Selection

Visualize and Explore

Compute Rank (Importance) of Nodes: Node Centralities & Advanced Graph Analytics / Machine Learning
O(N) methods for graph analytics: Stochastic approximation of matrix functions

Node importance: $F(A) = \expm(A)$

Graph comparison: $F(A) = \text{step function}$

Ineichen, Staar, Malossi, Bekas, Curioni, SC13, SC14, SC15
Dealing with complexity: The Cognitive Approach

• Combines the “collective knowledge” of the world with advanced simulation to create unprecedented progress

• Requires “old kind” of kernels and new data intensive workloads that are easily seen to surpass Exascale resources due the Data explosion and the need for ever increasing simulation fidelity

• New workloads come with non HPC traditional characteristics. Data Centric Systems and the flexibility of OpenPOWER are appropriate to tackle such challenges.
Toward the Forth Pillar of Innovation and Discovery

- Theory
- Simulations
- Data/Cognitive
- Experiment