Journey to Kitware

From experimental physics to scientific software development

Dr. Marcus D. Hanwell
@mhanwell
+MarcusHanwell
www.kitware.com

SUNY Buffalo, Buffalo, NY
1 April, 2016

Kitware, Inc.

Founded in 1998 by five former GE Research employees

107 current employees; 39 with PhDs

Privately held, profitable from creation, no debt

Offices

- Clifton Park, NY
- Carrboro, NC
- Santa Fe, NM
- Lyon, France

- 2011 Small Business Administration's Tibbetts Award
- HPCWire Readers and Editor's Choice
- Inc's 5000 List since 2008

Kitware's customers & collaborators

Over 75 **academic** institutions including…

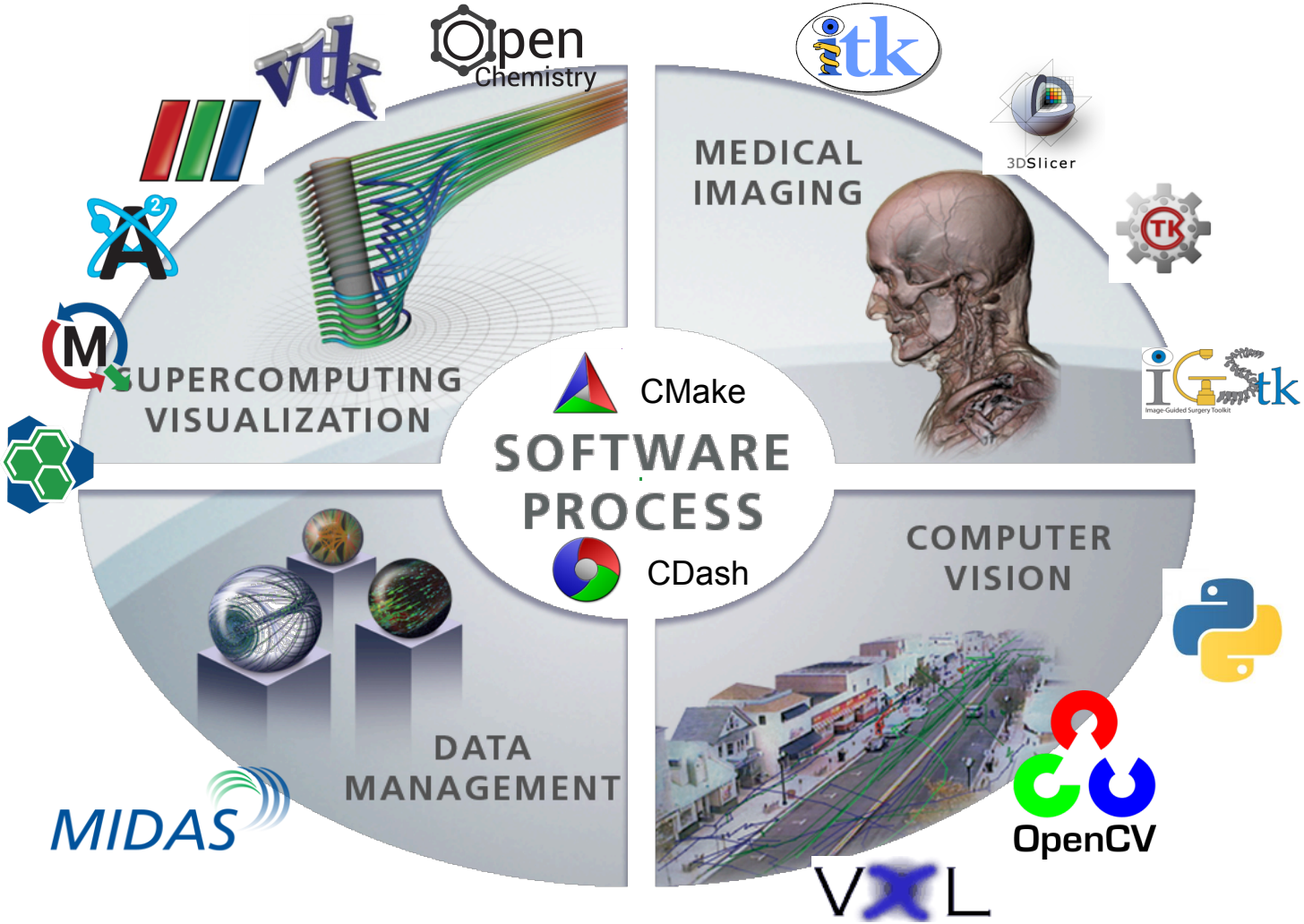
- Harvard
- Massachusetts Institute of Technology
- University of California, Berkeley
- Stanford University
- California Institute of Technology
- Imperial College London
- Johns Hopkins University
- Cornell University
- Columbia University
- Robarts Research Institute
- University of Pennsylvania
- Rensselaer Polytechnic Institute
- University of Utah
- University of North Carolina

Over 50 **government** agencies and labs including…

- National Institutes of Health (NIH)
- National Science Foundation (NSF)
- National Library of Medicine (NLM)
- Department of Defense (DOD)
- Department of Energy (DOE)
- Defense Advanced Research Projects Agency (DARPA)
- Army Research Lab (ARL)
- Air Force Research Lab (AFRL)
- Sandia (SNL)
- Los Alamos National Labs (LANL)
- Argonne (ANL)
- Oak Ridge (ORNL)
- Lawrence Livermore (LLNL)

Over 100 **commercial** companies in fields including…

- Automotive
- Aircraft
- Defense
- Energy technology
- Environmental sciences
- Finance
- Industrial inspection
- Oil & gas
- Pharmaceuticals
- Publishing
- 3D Mapping
- Medical devices
- Security
- Simulation


vtk
Open Chemistry
itk
3DSlicer
MEDICAL IMAGING
TK
IGSTK
Image-Guided Surgery Toolkit
SUPERCOMPUTING VISUALIZATION
CMake
SOFTWARE PROCESS
CDash
COMPUTER VISION
DATA MANAGEMENT
MIDAS
OpenCV
VXL

My Early Career

- 16 – Pizza delivery on a scooter
- 17 – Glass collector in a pub
- 17 – Left school, wasn’t sure it was for me
- 18 – Graduated to barman
- 19 – First software job, first headhunter
  - Later that year I quit and returned to school

Returning to School

- Careers service in my city helped a lot
- Enrolled for A-levels required for course
- Spoke to University of Sheffield
- After much deliberation – Physics
- Kept programming in my own time
- Regained enthusiasm for academia

Undergraduate Degree and Work

- Had to work part time
  - Got support from UK government for degree
- 20 – started Physics BSc
- 20 – part time at internet helpdesk
- 20 – Founded my own consultancy
- 21 – part time bar work off and on
- 22 – internship in Silicon Valley, CA, USA

Staying in School

- Originally intended to find work after degree
- Loved research, and wanted to do more
- Attained 1st class degree
- Offered fully funded PhD
- Growing Internet company asked me to interview in Silicon Valley that year too…

“Standing on the Shoulders of Giants”

Nullius in verba

Motto of the Royal Society, founded in 1660

“Take nobody’s word for it”

PHILOSOPHICAL TRANSACTIONS:

GIVING SOME

ACCOMPT

OF THE PRESENT

Undertakings, Studies, and Labours

OF THE

INGENIOUS

IN MANY

CONSIDERABLE PARTS

OF THE

WORLD.

Vol I.

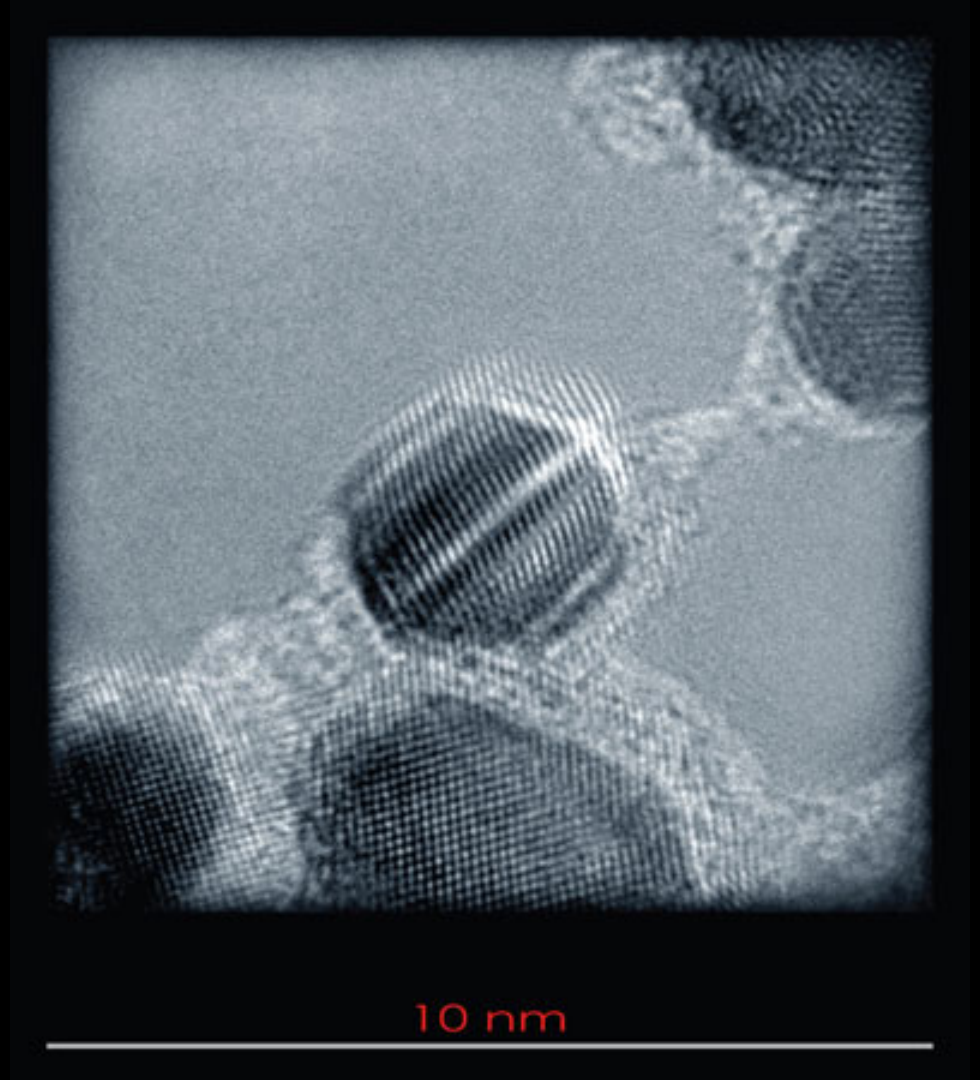
For *Anno* 1665, and 1666.

In the *SAVOY*,

Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*, Printers to the *Royal Society*.

PhD in Nanomaterial Engineering

- Experimental PhD in Physics
- Thiol encapsulated gold nanoparticles
- Langmuir-Blodgett films
- X-ray/neutron characterization
- 1-12 layer films deposited in clean room
- Electrical response to gas/vapours


10 nm

My Volunteer/Open Source Career

- 23 – became a Gentoo Linux developer
  - 64 bit porting and scientific applications
  - Learned Linux from building the kernel up
- 26 – started contributing to KDE
  - Wanted to actually create some packages
  - Big community, science sub-community
- Taught myself Bash, Python, C++, …

Running a Consultancy (20-27)

- Set up a one person company
- Worked for a number of local companies
- Setting up computers, installing Office
- Linux (in secret) as Internet gateway, firewall, email server
- Web hosting, web site design
- Security consulting, PHP, Perl, CGI

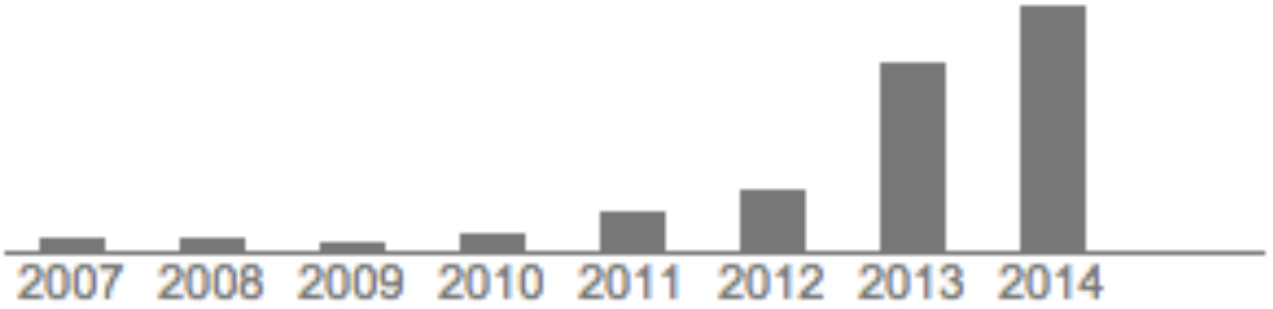
Google Summer of Code (27)

- Final year of my PhD – last chance to do it
- Supervisor agreed to three months off
- Found project in the KDE umbrella
  - Perfect fit – molecular editor for Kalzium
  - Contributed a few patches during application period
- Very productive summer, lots of progress
- Loved what I was doing

PhD and Postdoc (27) – Change

- Completed PhD, wasn’t sure what next
- Looking at being a dive guide for a year
- Offered a postdoc at Univ. of Pittsburgh
- Combining experiment with computation
  - Perfect fit, one of the founders of Avogadro
  - Experiment didn’t go well, equipment issues, students unable to synthesize compounds

Citation indices	All	Since 2009
Citations	423	394
h-index	7	7
i10-index	7	6



PHILOSOPHICAL

TRANSACTIONS:

GIVING SOME

ACCOMPT

OF THE PRESENT

Undertakings, Studies, and Labours

OF THE

INGENIOUS

IN MANY

CONSIDERABLE PARTS

OF THE

WORLD.

Vol I.

For *Anno* 1665, and 1666.

In the *SAVOY*,

Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,

Printers to the *Royal Society*.

Hanwell *et al. Journal of Cheminformatics* 2012, **4**:17
http://www.jcheminf.com/content/4/1/17

Journal of Cheminformatics

SOFTWARE **Open Access**

Avogadro: an advanced semantic chemical editor, visualization, and analysis platform

Marcus D Hanwell[1,2*], Donald E Curtis[3], David C Lonie[4], Tim Vandermeersch[5], Eva Zurek[4] and Geoffrey R Hutchison[1]

Abstract

Background: The Avogadro project has developed an advanced molecule editor and visualizer designed for cross-platform use in computational chemistry, molecular modeling, bioinformatics, materials science, and related areas. It offers flexible, high quality rendering, and a powerful plugin architecture. Typical uses include building molecular structures, formatting input files, and analyzing output of a wide variety of computational chemistry packages. By using the CML file format as its native document type, Avogadro seeks to enhance the semantic accessibility of chemical data types.

Results: The work presented here details the Avogadro library, which is a framework providing a code library and application programming interface (API) with three-dimensional visualization capabilities; and has direct applications to research and education in the fields of chemistry, physics, materials science, and biology. The Avogadro application provides a rich graphical interface using dynamically loaded plugins through the library itself. The application and library can each be extended by implementing a plugin module in C++ or Python to explore different visualization techniques, build/manipulate molecular structures, and interact with other programs. We describe some example extensions, one which uses a genetic algorithm to find stable crystal structures, and one which interfaces with the PackMol program to create packed, solvated structures for molecular dynamics simulations. The 1.0 release series of Avogadro is the main focus of the results discussed here.

Conclusions: Avogadro offers a semantic chemical builder and platform for visualization and analysis. For users, it offers an easy-to-use builder, integrated support for downloading from common databases such as PubChem and the Protein Data Bank, extracting chemical data from a wide variety of formats, including computational chemistry output, and native, semantic support for the CML file format. For developers, it can be easily extended via a powerful plugin mechanism to support new features in organic chemistry, inorganic complexes, drug design, materials, biomolecules, and simulations. Avogadro is freely available under an open-source license from http://avogadro.openmolecules.net.

Background

Many fields such as chemistry, materials science, physics, and biology, need efficient computer programs to both build and visualize molecular structures. The field of molecular graphics is dominated by viewers with little or no editing capabilities, such as RasMol [1], Jmol [2], PyMOL [3], VMD [4], QuteMol [5], BALLView [6], VESTA [7], and XCrySDen [8,9], among many others. The aforementioned viewers are all freely available, and most of them are available under open-source licenses and work on the most common operating systems (Linux, Apple Mac OS X, Microsoft Windows, and BSD).

The choice of software capable of building chemical structures in three dimensions is far smaller. There are existing commercial packages, such as CAChe/Scigress [10], ChemBio3D [11], GaussView [12], HyperChem [13], CrystalMaker [14], Materials Studio [15], and Spartan [16], which are polished and capable of constructing many different types of molecular structures. They are, however, not available for all operating systems (most of them only run on Microsoft Windows), and are not easily extensible, customized, or integrated into automated workflows.

*Correspondence: marcus.hanwell@kitware.com
[1]Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, PA, 15260, USA
[2]Department of Scientific Computing, Kitware, Inc., 28 Corporate Drive, Clifton Park, NY, 12065, USA
Full list of author information is available at the end of the article

Kitware (29)

- End of postdoc, at a crossroads
  - Gone from staying in academia to software
  - Wanted to keep doing science, open source
  - Looked at national labs, moving back to UK
- Gave a talk at Camp KDE in Jamaica
  - Mentioned I was starting to look for work
  - Several audience members seemed interested
  - Interviewed at Kitware a month later

Computer code is quickly replacing mathematics as the language of science

Kitware Career

- Hired as an R&D engineer
- Visas are hard, offer in Feb, start in Oct
- First few years working on charts, text processing, informatics
- Few years until I got some funding
- Promoted to Technical Lead (<3 years)

What Do I Do?

- Software development
  - C++, Python, JavaScript, HTML5, OpenGL, Java
- Software process
  - Windows, Mac, Linux, Git, code review, issues, …
- Proposals
  - SBIRs, collaborative, DOE, NSF, NIH, NIST, …
- Interviewing, recruiting, management, …

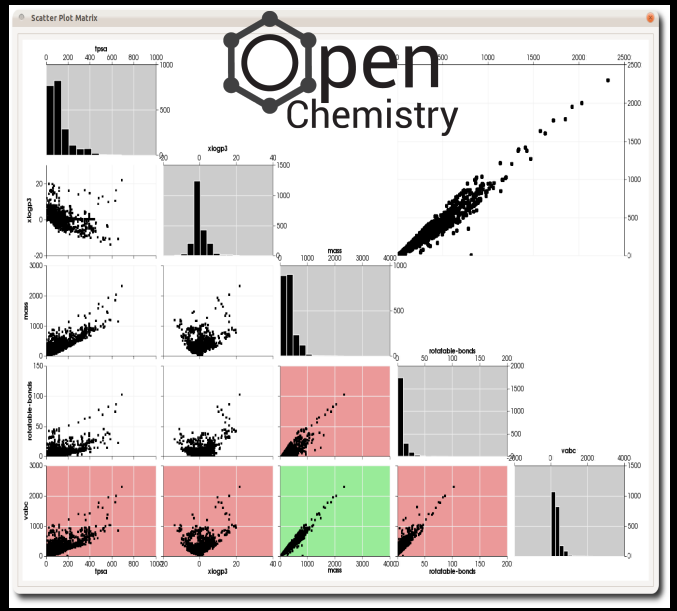
Awards and Grants

- ERDC SBIR Phase I & II – Open Chemistry
- DOE SBIR Phase I & II – Tomviz
- AFRL SBIR Phase I & II – Multiscale
- DOE SBIR Phase I, Phase II rejected
- Proposals rejected from NIH, NSF, DOE, DOD – tenacity is needed!
- Collaborative grants with different agencies

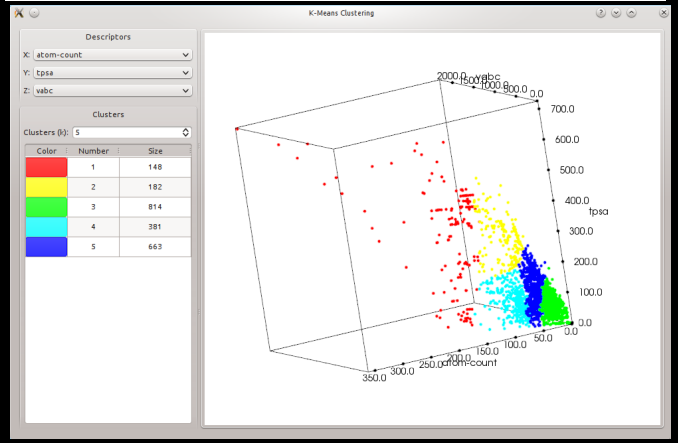
What Do We Look For?

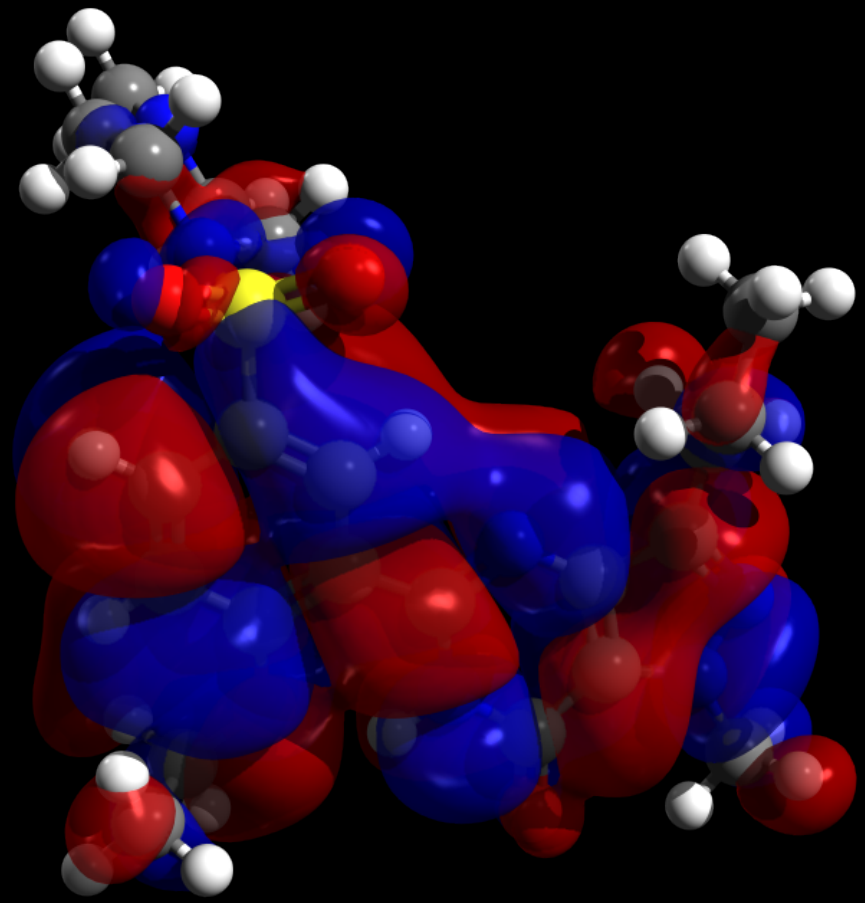
- Hire bachelors, masters, and doctorate
- Strong programming skills
  - C++, Python, HTML5, R
  - Windows, Mac, Linux, iOS, Android
- Software process, collaboration
  - Git, test driven development, code review
- Visualization, data processing, HPC, etc

Citation indices	All	Since 2011
Citations	884	830
h-index	9	8
i10-index	7	6



Open Chemistry





```
{
  "chemical json": 0,
  "name": "Ethane",
  "inchi": "1/C2H6/c1-2/h1-2H3",
  "formula": "C 2 H 6",
  "atoms": {
    "elements": {
      "number": [  1,   6,   1,   1,   6,   1,   1,   1 ]
    },
    "coords": {
      "3d": [  1.185080, -0.003838,  0.987524,
               0.751621, -0.022441, -0.020839,
               1.166929,  0.833015, -0.569312,
               1.115519, -0.932892, -0.514525,
              -0.751587,  0.022496,  0.020891,
              -1.166882, -0.833372,  0.568699,
              -1.115691,  0.932608,  0.515082,
              -1.184988,  0.004424, -0.987522 ]
    }
  },
  "bonds": {
```


$$GTO = ce^{-\alpha r^2}$$

$$\phi_i = \sum_{\mu} c_{\mu i}\phi_{\mu}$$

$$\rho(r) = \sum_{\mu}\sum_{\nu} P_{\mu\nu}\phi_{\mu}\phi_{\nu}$$

```
inline void GaussianSetTools::pointS(unsigned int moIndex, double dr2,
                                     vector<double> &values) const
{
  // S type orbitals - the simplest of the calculations with one component
  double tmp = 0.0;
  unsigned int cIndex = m_basis->cIndices()[moIndex];
  for (unsigned int i = m_basis->gtoIndices()[moIndex];
       i < m_basis->gtoIndices()[moIndex + 1]; ++i) {
    tmp += m_basis->gtoCN()[cIndex++] * exp(-m_basis->gtoA()[i] * dr2);
  }
  // There is one MO coefficient per S shell basis.
  values[m_basis->moIndices()[moIndex]] = tmp;
}

inline void GaussianSetTools::pointP(unsigned int moIndex, const Vector3 &delta,
                                     double dr2, vector<double> &values) const
{
  // P type orbitals have three components and each component has a different
  // independent MO weighting. Many things can be cached to save time though.
  unsigned int baseIndex = m_basis->moIndices()[moIndex];
  Vector3 components(Vector3::Zero());

  // Now iterate through the P type GTOs and sum their contributions
  unsigned int cIndex = m_basis->cIndices()[moIndex];
  for (unsigned int i = m_basis->gtoIndices()[moIndex];
       i < m_basis->gtoIndices()[moIndex + 1]; ++i) {
    double tmpGTO = exp(-m_basis->gtoA()[i] * dr2);
    for (unsigned int j = 0; j < 3; ++j) {
      //m_values[baseIndex + i] = m_basis->gtoCN()[cIndex++] * tmpGTO;
      components[j] += m_basis->gtoCN()[cIndex++] * tmpGTO;
    }
  }
  for (unsigned int i = 0; i < 3; ++i)
    values[baseIndex + i] = components[i] * delta[i];
}
```

```
{
  "chemical json": 0,
  "name": "Ethane",
  "inchi": "1/C2H6/c1-2/h1-2H3",
  "formula": "C 2 H 6",
  "atoms": {
    "elements": {
      "number": [ 1, 6, 1, 1, 6, 1, 1, 1 ]
    },
    "coords": {
      "3d": [ 1.185600, 0.000000, 0.987524,
              0.751087, 0.000000, 0.020839,
              1.166882, 0.833372, 0.569312,
              1.115519, -0.932892, -0.514525,
             -0.751587, 0.022496, 0.020891,
             -1.166882, -0.833372, 0.568699,
             -1.115691, 0.932608, 0.515082,
             -1.184988, 0.004424, -0.987522 ]
    }
  },
  "bonds": {
```


Data

$$GTO = ce^{-\alpha r^2}$$

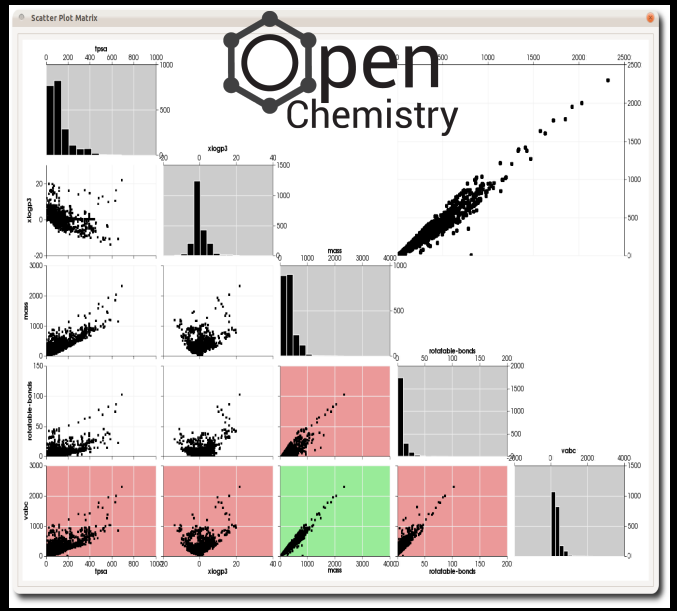
$$\phi = \sum_{\mu} c_{\mu i} \phi_{\mu}$$

Equations

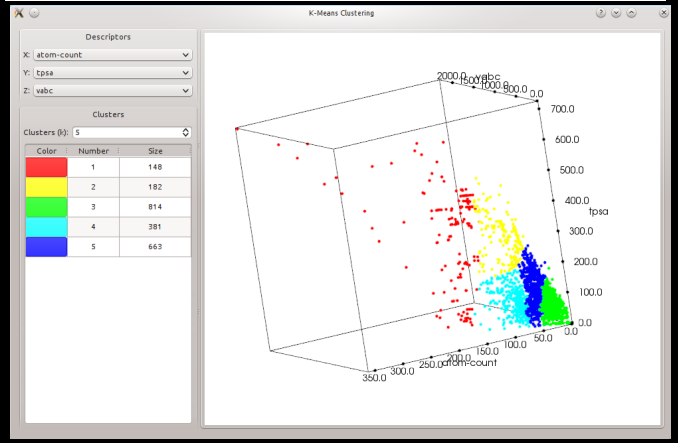
$$\rho(r) = \sum_{\mu} \sum_{\nu} P_{\mu\nu} \phi_{\mu} \phi_{\nu}$$

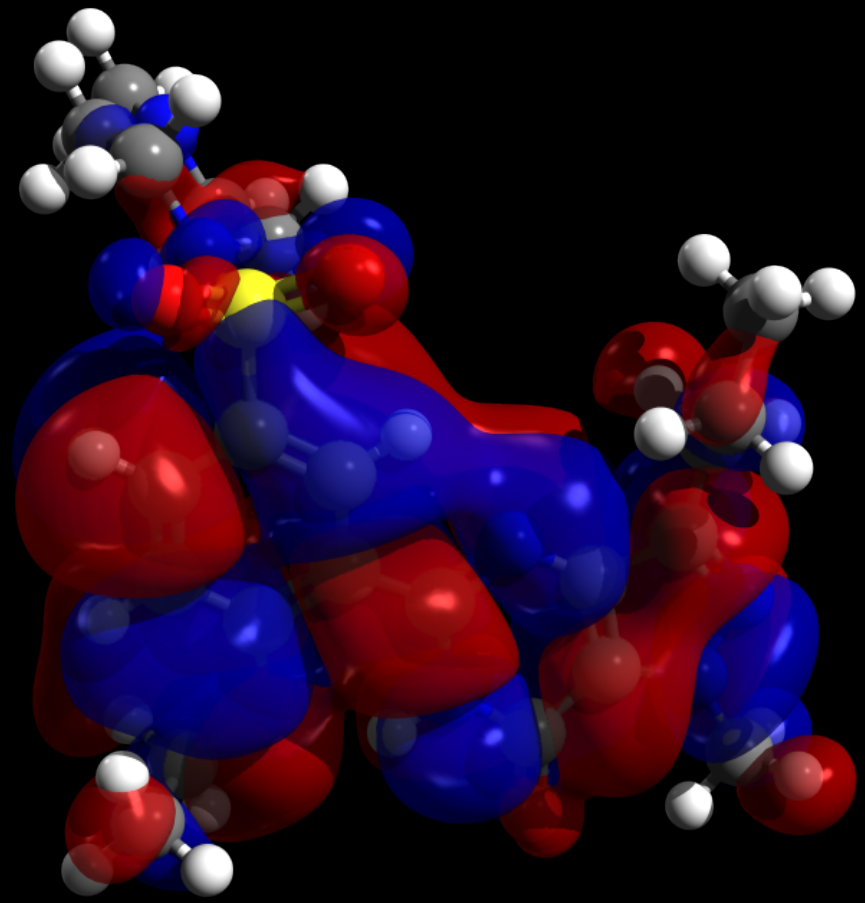
```
inline void GaussianSetTools::pointS(unsigned int moIndex, double dr2,
                                     vector<double> &values) const
{
  // S type orbitals - the simplest of the calculations with one component
  double tmp = 0.0;
  unsigned int cIndex = m_basis->cIndices()[moIndex];
  for (unsigned int i = m_basis->gtoIndices()[moIndex];
       i < m_basis->gtoIndices()[moIndex + 1]; ++i) {
    tmp += m_basis->gtoCN()[cIndex++] * exp(-m_basis->gtoA()[i] * dr2);
  }
  // There is one MO coefficient per S shell basis.
  values[m_basis->moIndices()[moIndex]] = tmp;
}

inline void GaussianSetTools::pointP(unsigned int moIndex, const Vector3 &delta,
                                     double dr2, vector<double> &values) const
{
  // P type orbitals have three components and each component has a different
  // independent MO weighting. Many things can be cached to save time though.
  unsigned int baseIndex = m_basis->moIndices()[moIndex];
  Vector3 components(Vector3::Zero());

  // Now iterate through the P type GTOs and sum their contributions
  unsigned int cIndex = m_basis->cIndices()[moIndex];
  for (unsigned int i = m_basis->gtoIndices()[moIndex];
       i < m_basis->gtoIndices()[moIndex + 1]; ++i) {
    double tmpGTO = exp(-m_basis->gtoA()[i] * dr2);
    for (unsigned int j = 0; j < 3; ++j) {
      //m_values[baseIndex + i] = m_basis->gtoCN()[cIndex++] * tmpGTO;
      components[j] += m_basis->gtoCN()[cIndex++] * tmpGTO;
    }
  }
  for (unsigned int i = 0; i < 3; ++i)
    values[baseIndex + i] = components[i] * delta[i];
}
```

Source Code



Open Chemistry





Do you see the mistake in the code?

How about now?

```
double Molecule::mass() const
{
  double result;
  for (Index i = 0; i < atomCount(); ++i)
    result += Elements::mass(atom(i).atomicNumber());
  return result;
}
```

Writing software for science the right way

Social Media

- LinkedIn for professional network
- Bought cryos.net domain
- Blogged very early (~2004)
- @mhanwell on Twitter
- +MarcusHanwell on Google+
- GitHub, Sourceforge, BitBucket, …
- Mixed on Facebook

Presentations

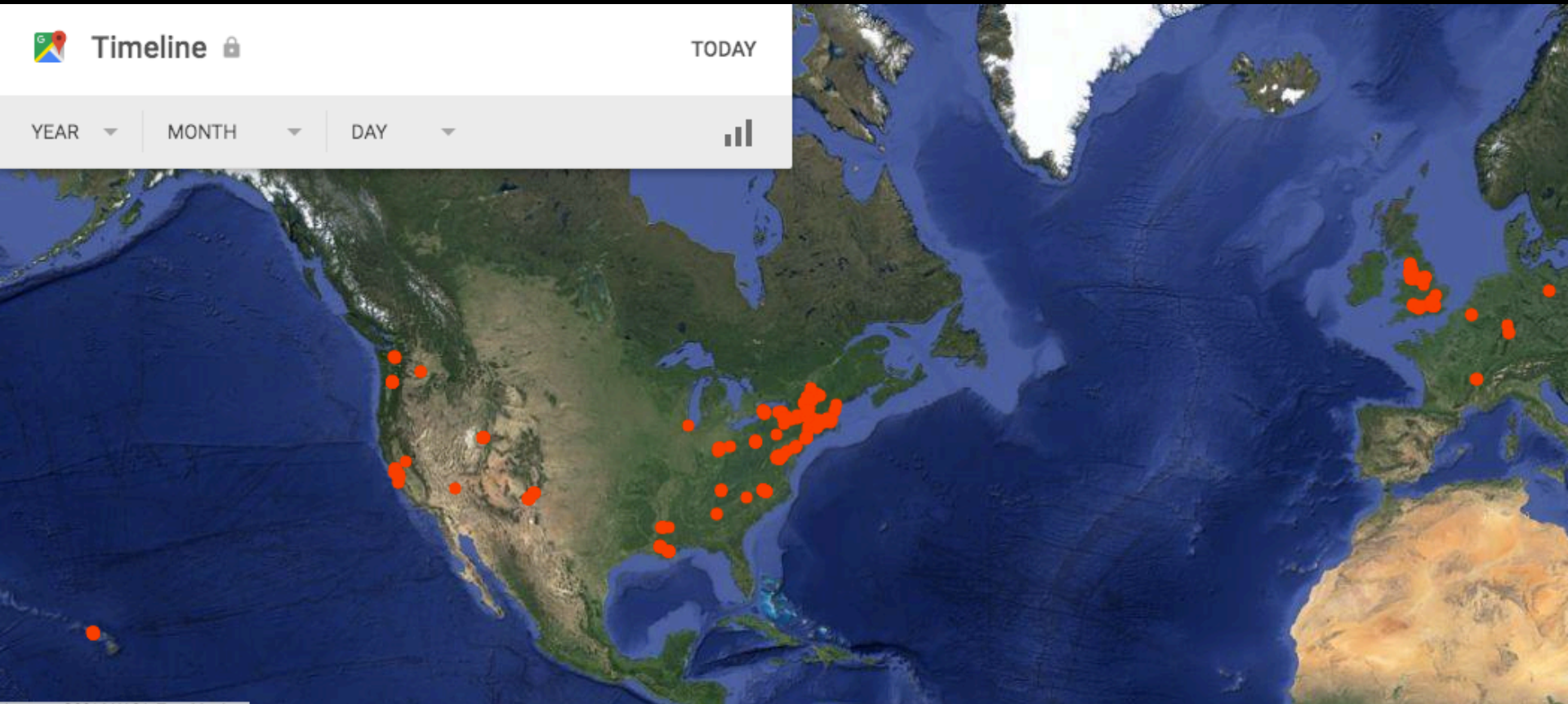
- Invited talks on science, open source, version control, software development
- ACS, APS, M&M, Langmuir, Pacifichem
- TEDxAlbany, OSCON, FOSDEM, All Things Open, Camp KDE, aKademy
- LBNL, LANL, ERDC, Sandia, SUNYs, LSU, RPI, NREL, Imperial, Harvard, Cornell

Passion

Skill

Communication

Travel



We are hiring!

http://jobs.kitware.com/