Leibniz: A Digital Scientific Notation

Konrad HINSEN

Centre de Biophysique Moléculaire, Orléans, France and Synchrotron SOLEIL, Saint Aubin, France

24 November 2016

Computational science today



Z. Merali, Nature 467, 775 (2010)

Konrad HINSEN (CBM/SOLEIL)

Leibniz: A Digital Scientific Notation

SCIENTIFIC PUBLISHING

A Scientist's Nightmare: Software Problem Leads to Five Retractions

Until recently, Geoffrey Chang's career was on a trajectory most young scientists only dream about. In 1999, at the age of 28, the protein crystallographer landed a faculty position at the prestigious Scripps Research Institute in San Diego, California. The next year, in a ceremony at the White House, Chang received a 2001 Science paper, which described the structure of a protein called MsbA, isolated from the bacterium *Escherichia coli*. MsbA belongs to a huge and ancient family of molecules that use energy from adenosine triphosphate to transport molecules across cell membranes. These so-called ABC transporters perform many

G. Miller, Science 314 1856 (2007)

OPEN ORCESS Freely available online



The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements

Ed H. B. M. Gronenschild^{1,2}*, Petra Habets^{1,2}, Heidi I. L. Jacobs^{1,2,3}, Ron Mengelers^{1,2}, Nico Rozendaal^{1,2}, Jim van Os^{1,2,4}, Machteld Marcelis^{1,2}

E. Gronenschild et al., PLoS ONE 7(6) e38234 (2012)

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund^{a,b,c,1}, Thomas E. Nichols^{d,e}, and Hans Knutsson^{a,c}

^{*}Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweder, ^{*}Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweder, ^{*}Center for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweder, ^{*}Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom; and *WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

A. Eklund, T.E. Nichols, H. Knutsson, PNAS 113(28) 7900 (2016)

PNAS

- Insufficient software engineering
- Testing numerical software is very difficult.
- Most scientific software has no specification.
- Most scientific software is badly documented.

- Reliable tests check for equality...
- ... but exact equality never happens with floats.
- ... but exact equality never happens with floats. Why ???
- ... therefore we must introduce tolerances ...
- ... but nobody knows how to choose them.

Full discussion:

K. Hinsen, The Approximation Tower in Computational Science: Why Testing Scientific Software Is Difficult, Comp. Sci. Eng. **17**(4) 72-77 (2015)

Informal specifications / documentation

- Theory: should be published in journal articles etc.
- Practice: complexity of models and methods prevents a full description

Formal specifications

- Unknown to most computational scientists
- No domain-specific languages and tools

Climate models

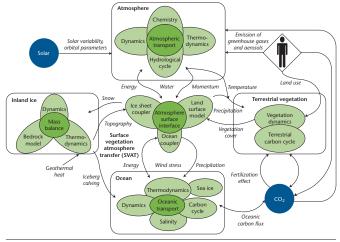


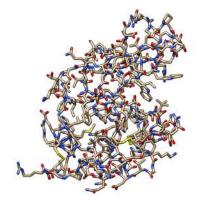
Figure 1. Conceptual view of the components and couplings of a coupled Earth system model.

From: Easterbrook & Johns, Comp. Sci. Eng. 11(6), 65-74 (2009)

Konrad HINSEN (CBM/SOLEIL)

Leibniz: A Digital Scientific Notation

A small protein (lysozyme) in the Amber99 force field



- 1960 atoms (1001 shown)
- 183 distinct atom types
- interaction energy: a function of 5880 variables, with a few thousand numerical parameters

Remember that this is a **small** protein!

Biomolecular force fields

(1/2)

This is how a typical research paper describes the AMBER force field:

$$U = \sum_{\text{bonds } ij} k_{ij} \left(r_{ij} - r_{ij}^{(0)} \right)^{2}$$

+
$$\sum_{\text{angles } ijk} k_{ijk} \left(\phi_{ijk} - \phi_{ijk}^{(0)} \right)^{2}$$

+
$$\sum_{\text{dihedrals } ijkl} k_{ijkl} \cos \left(n_{ijkl} \theta_{ijkl} - \delta_{ijkl} \right)$$

+
$$\sum_{\text{all pairs } ij} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r^{12}} - \frac{\sigma_{ij}^{6}}{r^{6}} \right)$$

+
$$\sum_{\text{all pairs } ij} \frac{q_{i}q_{j}}{4\pi\epsilon_{0}r_{ij}}$$

Konrad HINSEN (CBM/SOLEIL)

Protein parameter files for Amber 12, in somewhat documented formats:

lines	filename
984	amino12.in
814	aminoct12.in
782	aminont12.in
533	frcmod.ff12SB
744	parm99.dat

For the details... read the source code!

For the algorithms that select the parameters for a given protein structure... read the source code!

SOFTWARE ENGINEERING FOR CSE

Streamlining Development of a Multimillion-Line Computational Chemistry Code

Robin M. Betz and Ross C. Walker I San Diego Supercomputer Center

Software engineering methodologies can be helpful in computational science and engineering projects. Here, a continuous integration software engineering strategy is applied to a multimillion-line molecular dynamics code; the implementation both streamlines the development and release process and unifies a team of widely distributed, academic developers.

Betz & Walker, Comp. Sci. Eng. 16(3), 10-17 (2014)

Konrad HINSEN (CBM/SOLEIL)

Science requires **verifiability** of all findings, but:

- We cannot discuss computational models and methods ... because we cannot write them down.
- We cannot compare different models and methods ... for the same reason.
- We cannot test scientific software ... because we have no specifications.
- Users don't understand their software
 - ... because it's too complex ...
 - ... and there is no specification.

Reports and Articles

Social Processes and Proofs of Theorems and Programs

Richard A. De Millo Georgia Institute of Technology

Richard J. Lipton and Alan J. Perlis Yale University

Communications of the ACM, 1979

Similar issues elsewhere (2/2)

In Computers We Trust?

As math grows ever more complex, will computers reign?



Illustration by Simons Science News

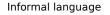
This simple computation, written with math software called Maple, verifies a formula for the number of integer triangles with a given perimeter.

By Natalie Wolchover

Quanta Magazine, 2013

Konrad HINSEN (CBM/SOLEIL)

Leibniz: A Digital Scientific Notation



THE R. LEWIS CO., LANSING MICH.

the deal artistation have rather with their month in control collector observes of exactlying blacks. The complete positions dealers in general that price area to considere difference dealers article article and price area for considere difference dealers are dealers of positions. The producting interface legislation of parts

PROPERTY AND ADDRESS

we be composite of the constraint and in 5 of the crosses. For $\mu < 1$, 2007 orders in market, there exists the second constraints of the constraint of th

et source lagrande. Experim ensues has not de la la classificación en la concentra de la classificación de la concentra en la concentra de la cuasión de la concentra en la concentra de la cuasión de la concentra de la concentra de la concentra de la concentra de la cuasión de la concentra de la concentra de la concentra de la cuasión de la concentra de la concentra de la cuasión de la concentra de la concentra de la concentra de la cuasión de la concentra della concentra della concentra de la concentra della concentra della concentra de la concentra della concentra

Antonios Tra uma preto el fuquence alternere en a prepladade atrabación de la del activación de la prepladade atrabación de la del activación de la del alternación per la del con de pracamente de la del de deservación, ello esta indexidade de la del de deservación, ello esta atrabación de la del de deservación de la cuente atrabación de la del de deservación de la del de del de del de del de deservación de la del de del de del de del de deservación de la del de del de del de del de deservación de la del de del de del de del deservación de del deservación de la del de deservación de del deservación de del de del de deservación de deservación de del de del de del deservación de deservación del de deservación de del deservación de deservación del de deservación de del deservación de deservación de del deservación de des

 $\lambda_{\rm constant} = \left(\left(\frac{1}{1+1} \frac{\partial |V_{\rm constant}|}{\partial |V_{\rm constant}|} + \partial |V_{\rm constant}| \right) \right) = 0$

eventing bit for transition are transition unlimited of the BM assume. The factor (UU) is U_{ij} is able to implement the subscription PU are a part of U_{ij} (2). It includes the international actor is an L_{ij} and U_{ij} (2). It includes the internation actor is an L_{ij} are implemented in the operation. The international T_{ij} (4) is a strateging internet operation.

Support relative for several attempts out to malacanet appendix to adopt place process for out a path attempt of the property of addition of the self-several attempts of the property of addition. It is not adopt of the property of addition indicates of the path; of addition and being the indicates of the path; of addition and of a of appendix of the path; of addition and of a of a property of the path; of addition and a set of an appendix of the path; of addition and a set of an 1000.00100.000

durant for space box. Not, it unless all the strangtone due to an ensure the lowest models. Without the second strategies and the strategiest of the gamma second strategiest strategiest possibility distribution from Fig. 30, its coupled this fundmental distribution from Fig. 30, its coupled this fundmental distribution from Fig. 30, its coupled this fundmental distribution for the strategiest strategiest possibility distribution for the strategiest strategiest of the strategiest distribution of the strategiest strategiest strategiest of the strategiest distribution of the strategiest strategiest strategiest of the strategiest distribution of the strategiest strategiest strategiest of the strategiest strategiest strategiest strategiest strategiest strategiest of the strategiest strategiest strategiest strategiest strategiest strategiest strategiest strategiest of the strategiest strategiest

$$\begin{split} & \| \mathcal{A}_{n}(x) ^{*} |_{X} + \left(\mathcal{A}_{n,1}^{(1)} + \mathcal{L}^{n}(\mathbb{F}_{n,1}^{(1)}) \right) & \text{if } \\ & \| \mathbf{A}_{n} \|_{H} & \| \mathbf{A}_{n}(x) ^{*} \|_{H} + \| \mathbf{A}_{n}(x$$

Before typictory large an electronic to the eccentration of this parameters in the district a parate spice spice spice spice spice for the parameters in the spice of the parameters of the spice spice spice is a spice of the spice of the spice spice spice spice spice spice spice spice spice of the spice spice spice spice spice spice spice spice spice of the spice spice spice spice spice spice spice spice spice of the spice spice

For the set of the perturbation of the set of the set

The case amplitude while subjective is to an and property entermode the compared is the second-size-offset by the DM strends, buggethade the functionary where the derivative of the caseful to generate a generative boots works, is with transition the support of other strends and of caseful to strength and the support of the strends and of caseful to strength and the strength and the DM strend the two derives are strength to the DM strend the two derives are strength to the DM strend the two derives are strength to the DM strend the strength of the strength to the DM strend the two derives are strength to the DM strend the two derives are strength to the DM strend the two derives are strength to the DM strend to

To enclose our plane, an enclose to relation of this particular planets planet, E.C., - A. Tolomannam in a classical planets planet is set, which is also and relation framework planets by considered length 2. We by: ETCL

$$\begin{split} & \left\{ \sum_{i=1}^{n} < d_{i,i} d^{2} \left(||\mathbf{x}| + (||^{2} - 2||^{2} + ||\mathbf{x}| - ||^{2} \right) ||\mathbf{x}| + \mathbf{x} + \right\} \\ & \text{To note framewer process, } \Sigma_{i} \text{ and } \sum_{i=1}^{n} d_{i,i} d^{2} \left(\sum_{i=1}^{n} d_{i,i} d^{2} \right) \right\} \end{split}$$

Digital Scientific Notation

formal • embedded into

scholarly discourse

automated (partial) validation

verifiable by human readers

programming language

intervention of the second section of the first second secon

Essay and Opinion

Scientific notations for the digital era

VERSION 1 Released on 27 April 2016 under Creative Commons Attribution 4.0 International License

Konrad Hinsen^{1,2}

Authors' affiliations

1. Centre de biophysique moléculaire (CBM). CNRS : UPR4301.

2. Synchrotron SOLEIL,

K. Hinsen, The Self-Journal of Science, 2016

Konrad HINSEN (CBM/SOLEIL)

thesis

Digital science

Technology is transforming science. We can calculate faster than ever before, and gather more precise data, of an everexpanding variety. We now store staggering volumes of information, and draw on it ever more creatively. Communication between scientists is changing too, even if the scientific paper still remains the standard element of information transfer and recording. The web, e-mail and blogs, as well as social media including Facebook and



apparatus was set up and used; we accept theoretical arguments only if each step is

reduced to numerics. The result is a series of diabolical tendencies all too familiar to computational scientists. As he notes, scientists can develop software using the best practices of software engineering and the result may still compute something different from what its users think it computes. Of course, concerned users can in principle consult the documentation, but how do they know that documentation is complete and accurate?

M. Buchanan, Nature Physics 12 630 (2016)

- My first attempt at a Digital Scientific Notation
- Still in a very early stage
- Any feedback welcome:
 - Open Science project on Digital Scientific Notations in general
 - Leibniz on GitHub

Gottfried Wilhelm Leibniz (1646-1716)

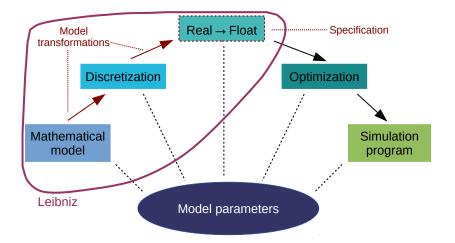


Leibniz attributed all his discoveries in mathematics to the development and use of good notations.

Ideas most relevant to Digital Scientific Notations:

- Leibniz's notation in calculus: dy/dx, \int etc.
- Calculus ratiocinator: computing, formal logic, etc.

Approximation chain in numerical simulations



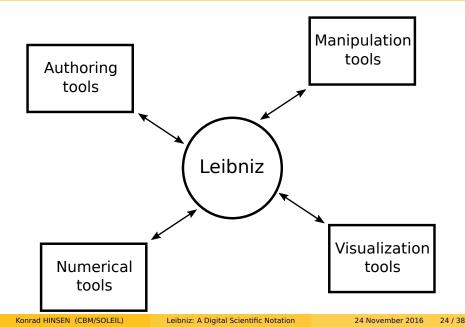
Konrad HINSEN (CBM/SOLEIL)

Leibniz syntax

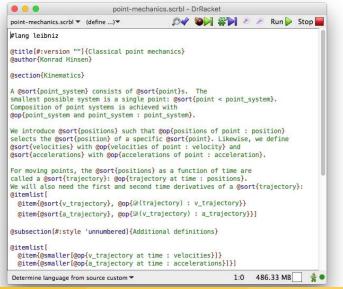
```
<context id="point-kinematics">
  <sorts>
   <sort id="point_system" />
   <sort id="velocities" />
   <sort id="a_trajectory" />
   <sort id="positions" />
   <sort id="velocity" />
   <sort id="position" />
   <sort id="v_trajectory" />
   <sort id="accelerations" />
   <sort id="acceleration" />
   <sort id="trajectory" />
   <sort id="point" />
   <sort id="time" />
  </sorts>
  <subsorts>
   <subsort sort="point" sort="point_system" />
  </subsorts>
  <0DS>
   <op id="at">
     <arity>
       <sort id="a_trajectory" />
       <sort id="time" />
      </arity>
      <sort id="accelerations" />
   </00>
```

. . .

Software tools around Leibniz



Leibniz authoring environment (just a demo for now)



Konrad HINSEN (CBM/SOLEIL)

Leibniz: A Digital Scientific Notation

Output for human readers

Classical point mechanics

1 Kinematics

A *point_system* consists of *points*. The smallest possible system is a single point: *point* ⊆ *point_system*. Composition of point systems is achieved with *point_system* and *point_system*.

We introduce *positions* such that *positions* of *point*: *position* selects the *position* of a specific *point*. Likewise, we define *velocities* with *velocities* of *point*: *velocity* and *accelerations* with *accelerations* of *point*: *acceleration*.

For moving points, the *positions* as a function of time are called a *trajectory*: *trajectory* at *time*: *positions*. We will also need the first and second time derivatives of a *trajectory*:

- v_trajectory, D(trajectory): v_trajectory
- a_trajectory, D(v_trajectory) : a_trajectory

1.1 Additional definitions

"Development" syntax (works now)

lang sweet-exp racket	Alang racket
provide (all-defined-out)	(provide mass mass-units)
require leibniz */quantities/mass.rkt*	(reguire léibhiz)
"/quentitles/space.rkt"	(module+ test
"/quantities/time.rkt"	(require rackunit))
module+ test	(define-context mass
require rackunit	
	(include real-numbers)
	(sort Hass)
lefine-context point-mass-system	(op (Mass , + , Mass) Mass)
	(op (PositiveReal . * . Mass) Mass)
Include mass	(op (Mass , / , PositiveReal) Mass)
	(op (Mess , / , Mass) PositiveReal)
The sort of point mass identifiers.	
sort PointMass	(⇒ #:vars ([#1 PosttiveReal] [#2 PosttiveReal] [M Mass])
	(F1.*.(F2.*.M))
sort PointMassSystem	((F1 , * , F2) , * , N))
	(🖙 #:vars.([P PosttiveReal] [M Hass])
subsort PointMass PointMassSystem	
	((1.7.F),*.N))
sort Masses	(⇒ #:vers ([# PositiveReal] [M1 Mass] [M2 Mass])
	<pre>> (M1 _ / . (* F HZ))</pre>
	((/H1F)_/_H2))
op empty-space PointMassSystem	(=> #rvars ([F PositiveReal] [M1 Mass] [M2 Mass])
; Any combination of point-mass systems is a point-mass system	((F.*.Hl)./.M2)
op (PointMassSystem and PointMassSystem) PointMassSystem	(F. * , (/ H1 H2)))
	(+> #:var (M Hass)
: A selection operator that extracts one mass from the	(M., Z., H)
collection of all masses. op (Masses of PointMass) Mass	
op (Masses of Politinass) hass	(define-context mass-units-
The name of the collection of masses	(define-context mess-units (include mess)
op wass Masses	(Include Mass) (Include boolean)
op mass masses /: point-mechanics.rkt Top L2 Git-demo-preparation (Racket hs	

- An algebraic specification language
- based on equational logic
- ... and term rewriting

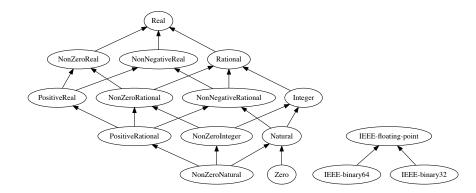
Today, Leibniz is a subset of Maude with different syntax.

In the future, Leibniz will further diverge from Maude.

Order-sorted term algebra

- A Leibniz term is either a standard term or a special term.
- A standard term has the form *op*(*arg*1 *arg*2 ...), or just *op* for zero arguments.
- Special terms are integers, rationals, and floats.
- Each term has an associated sort.
- Sorts have a partial order, defined by subsort relations.
- Subsort relations form a directed acyclic graph.
- The connected components of this graph are called kinds.
- All terms must be kind-correct ("static typing")
- Sort errors within a kind can be resolved during rewriting ("dynamic typing")

Sort graph for the number sorts



```
define-context mass
  include real-numbers
  sort Mass
  .
  ,
  : The sum of two masses is a mass.
  op {Mass + Mass} Mass
  ; The product of a positive number with a mass is a mass.
  op {PositiveReal * Mass} Mass
  ; A mass divided by a positive number is a mass.
  op {Mass / PositiveReal} Mass
  ; The quotient of two masses is a positive number.
  op {Mass / Mass} PositiveReal
```

```
define-context distance
  include real-numbers
  ;
  sort Distance
  ,
  sort NonZeroDistance ; an important special case for division
  subsort NonZeroDistance Distance
  ;
 op {Distance + Distance} Distance
  op {Distance - Distance} Distance
  ;
  op {Real * Distance} Distance
  op {NonZeroReal * NonZeroDistance} NonZeroDistance
  op {Distance / NonZeroReal} Distance
  ;
  op {Distance / NonZeroDistance} Real
  op {NonZeroDistance / NonZeroDistance} NonZeroReal
```

Terms

```
define-context distance-test
  include distance
  op d Distance
  op nzd NonZeroDistance
with-context distance-test
  displayln T(d)
  displayln T(nzd)
  displayln T{2 * d}
  displayln T{2 * nzd}
  displayln T{0 * nzd}
  displayln T{d / nzd}
  displayln T{nzd / d}
```

```
Distance:d
NonZeroDistance:nzd
Distance:(* 2 d)
NonZeroDistance:(* 2 nzd)
Distance:(* 0 nzd)
Real:(/ d nzd)
[NonZeroNatural]:(/ nzd d)
```

Rewrite rules

```
define-context distance-with-rules
  .
  ;
  include distance
  ;
  => ∀ D : Distance
     \{1 * D\}
     D
  => ∀ F1 : Real
     ∀ F2 : Real
     ∀ D : Distance
     \{\{F1 * D\} + \{F2 * D\}\}
     \{\{F1 + F2\} * D\}
  => ∀ F : NonZeroReal
     ∀ D : Distance
     {D / F}
     \{\{1 / F\} * D\}
  => ∀ F1 · Real
     ∀ F2 · Real
     ∀D : Distance
     {F1 * {F2 * D}}
     {{F1 * F2} * D}
```

define-context distance-test
 include distance-with-rules
 op d Distance

with-context distance-test
displayIn T{{2 * d} + {3 * d}}
displayIn RT{{2 * d} + {3 * d}}
displayIn RT{{2 * d} + {d / 2}}

```
Distance:(+ (* 2 d) (* 3 d))
Distance:(* 5 d)
Distance:(* 5/2 d)
```

Equations and transformations

```
define-context distance-with-equations
  ;
  include distance-with-rules
  ;
  op d1 Distance
  op d2 NonZeroDistance
  ;
  eq #:label eq-1
     d1
     \{2 * d2\}
with-context distance-with-equations
  ;
  displayln eq(eq-1)
                                         (eq #:label eq-1 d1 (* 2 d2))
  displayln
                                         (eg #:label eg-2 (* 1/2 d1) d2)
    А
      tr #:var (D Distance) D {D / 2}
      eq(eq-1)
      #:label eq-2
```

Embedding code into scholarly discourse leads to different priorities from software development.

- No namespaces, no scopes, no modularity.
- Use explicit renaming instead (supported by the authoring environment).
- Minimal built-in contexts: just numbers and booleans.
- No "standard library".
- Re-use and adapt published libraries instead.
- Prevents the creation of large black-box code libraries.

- All the code is on GitHub.
- Written in Racket, which provides excellent support for this kind of project.

Future work

Leibniz

- Associative/commutative operators
- Built-in collections: lists, sets, maybe more
- Interfaces to external data (files, databases, ...)

Support tools

- Authoring environment
- Manipulation tools for numerical work (real ightarrow float etc.)
- Libraries for popular programming languages

Find funding

- Project outside of every funding agency's categories
- Suggestions welcome!