Refactoring, reengineering and evolution: paths to Geant4 uncertainty quantification and performance improvement

Maria Grazia Pia
M. Batič, M. Begalli, M. Han, S. Hauf, G. Hoff, C. H. Kim, M. Kuster, P. Saracco, H. Seo, G. Weidenspointner, A. Zoglauer

INFN Genova, Italy
State University of Rio de Janeiro (UERJ), Brazil
Hanyang University, Seoul, Korea
Darmstadt Tech. Univ., Germany
PUCRS, Porto Alegre, Brazil
EU XFEL GmbH, Hamburg, Germany
MPI Halbleiterlabor, Munich, Germany
Space Sciences Laboratory, UC Berkeley, USA
“The code slowly sinks from engineering to hacking”.

“Refactoring is the process of changing a software system in such a way that it does not alter the external behavior of the code yet improves its internal structure.”

“When you refactor you are improving the design of the code after it has been written.”

Reengineering “seeks to transform a legacy system into the system you would have built if you had the luxury of hindsight and could have known all the new requirements that you know today.”

“A legacy is something valuable that you have inherited.”

“With rapid development tools and rapid turnover in personnel, software systems can turn into legacies more quickly than you might imagine.”

Maria Grazia Pia, INFN Genova
1. one of a set of prescribed movements
2. a process of change in a certain direction
3. the process of working out or developing
4. the historical development of a biological group
5. the extraction of a mathematical root
6. a process in which the whole universe is a progression of interrelated phenomena
The lantern

State of the art
in the physics for Monte Carlo particle transport

Quantified simulation

Focus on **physics** and **fundamental concepts**
of particle transport

Theoretical and technological challenges

Produce concrete deliverables

Maria Grazia Pia, *INFN Genova*
State-of-the-art, quantified

Geant 4

R&D Project
Series of pilot projects going on since 2008

Highlights
No time to go into details

Archival literature
Maria Grazia Pia, INFN Genova
Smells

If it stinks, change it.
Grandma Beck, discussing child-rearing philosophy

M. Fowler, K. Beck et al.,
Refactoring: Improving the Design of Existing Code

- Duplicated Code
- Long Method
- Large Class
- Long Parameter List
- Divergent Change
- Shotgun Surgery
- Feature Envy
- Data Clumps
- Primitive Obsession
- Switch Statements
- Parallel Inheritance Hierarchies

- Lazy Class
- Speculative Generality
- Temporary Field
- Message Chains
- Middle Man
- Inappropriate Intimacy
- Alternative Classes with Different Interfaces
- Incomplete Library Class
- Refused Bequest

Maria Grazia Pia, INFN Genova

Grandma Beck, discussing child-rearing philosophy

M. Fowler, K. Beck et al.,
Refactoring: Improving the Design of Existing Code

- Duplicated Code
- Long Method
- Large Class
- Long Parameter List
- Divergent Change
- Shotgun Surgery
- Feature Envy
- Data Clumps
- Primitive Obsession
- Switch Statements
- Parallel Inheritance Hierarchies

- Lazy Class
- Speculative Generality
- Temporary Field
- Message Chains
- Middle Man
- Inappropriate Intimacy
- Alternative Classes with Different Interfaces
- Incomplete Library Class
- Refused Bequest

Maria Grazia Pia, INFN Genova
Electromagnetic smells

**Coupling**

$\sigma_{\text{tot}}$ and final state modeling have been decoupled in hadronic physics design since RD44

**Dependencies**

on other parts of the software

One needs a geometry (and a full scale application) to test (verify) a cross section

Difficult to test $\Rightarrow$ no testing often

**“model”**

- Total cross section
- Whether a process occurs
- Final state generation
- How a process occurs

**Problem domain analysis**

Improve domain decomposition

Maria Grazia Pia, INFN Genova
Policy-based class design

Benefits

Transparency
Ease of maintenance
Simplicity of testing for V&V

Numera ciò che è numerabile, misura ciò che è misurabile,
e ciò che non è misurabile rendilo misurabile.
Galileo Galilei (1564-1642)

Basic physics V&V can be performed by means of lightweight unit tests
Exploring new models (calculations) is made easier
Quantification of accuracy is facilitated

Maria Grazia Pia, INFN Genova
QED ≠ QCD

Electromagnetic processes in particle transport:
- Final state to be generated is well identified
- **Theory**, rather than **model**
- Various degrees of refinement in theoretical calculations
  - e.g. electron at rest, scattering functions, Compton profiles
- Experimental data for validation are (in general) available

Bare-bone physical functionality

Decorations on top

<table>
<thead>
<tr>
<th>Compton scattering</th>
<th>Target electron at rest</th>
<th>Recoil electron</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident photon</td>
<td>Scattered photon</td>
<td></td>
</tr>
</tbody>
</table>

Motion of atomic electrons
Doppler broadening

Vacancy creation
Atomic relaxation

Binding effects

scattering functions

Doppler profiles

Distinguish genuine modeling alternatives from evolving degree of complexity

Shift from shopping list of alternatives

Maria Grazia Pia, INFN Genova
An example:
Photon elastic scattering

State of the art
Form factor approximation:
non relativistic, relativistic, modified + anomalous scattering factors

2\textsuperscript{nd} order S-matrix calculations
recent calculations, not yet used in Monte Carlo codes

Quantification
Statistical analysis, GoF + categorical

Differential cross sections

<table>
<thead>
<tr>
<th></th>
<th>Penelope Penelope</th>
<th>EPDL</th>
<th>Relativ. FF</th>
<th>Non-Rel. FF</th>
<th>Modified FF</th>
<th>MFF ASF</th>
<th>RFF ASF</th>
<th>SM NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.27</td>
<td>0.38</td>
<td>0.38</td>
<td>0.25</td>
<td>0.35</td>
<td>0.49</td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td>2008</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.25</td>
<td>0.35</td>
<td>0.49</td>
<td>0.52</td>
<td>0.77</td>
</tr>
</tbody>
</table>

\(\varepsilon\) = fraction of test cases compatible with experiment, 0.01 significance
Prune

Number one in the stink parade is duplicated code physics

Objective quantification of smell

Two Geant4 models, identical underlying physics content (it used to be different)

Efficiency w.r.t. experiment

“Livermore” | Penelope
EPDL97      | EPDL97
0.38±0.06   | 0.38±0.06

Code bloat
Burden on
• Software design
• Maintenance
• User support
Unnecessary complexity

Bremsstrahlung, evaporation, proton elastic scattering etc.
Atomic binding energies

Geant 4: Carlson + Williams, EADL (Carlson + Shirley)

Source of epistemic uncertainties?

Evaluation of Atomic Electron Binding Energies for Monte Carlo Particle Transport

Maria Grazia Pia, Hee Seo, Matej Batic, Marcia Begalli, Chan Hyeong Kim, Lina Quintieri, and Paolo Saracco
Algorithms

Popular belief

Physics model X is intrinsically slow

Baroque methods to combine it with “faster” lower precision models and limit its use to cases where one is willing to pay for higher precision

This design introduces an additional computational burden due to the effects of inheritance and the combination algorithms themselves

Truth

Physics model X is intrinsically fast

But its computationally fast physics functionality is spoiled by an inefficient sampling algorithm

No code smell

Spotted through in-depth code review in the course of software validation

Change the sampling algorithm!
The fastest algorithm

no algorithm at all

Shift modeling from algorithms to data
### Merging models

**Smoothing data**

- Guidance from experimental data (when available)

---

**Electron impact ionisation cross sections**

![Graph showing electron impact ionisation cross sections for Mg (Z=12).](image-url)

- **Total cross section (Mb)**
- **Energy (MeV)**

Example: LOESS local polynomial regression fitting

Beware: not optimized! Mathematical support in progress
...no silver bullet

Smoothing models (or data) that exhibit significantly different behaviour is not physically justified

Here our understanding of the underlying physics phenomena fails

Experimental data themselves must be analyzed for inconsistencies and possible systematics

How to solve this kind of problems?

Better theoretical calculations
New, dedicated experimental data

No easy solution

Photon elastic scattering cross section at 90°

Experimental data by Schumacher et al., Starr et al., Jackson et al., Moreh et al. etc., complete list in a forthcoming publication

Maria Grazia Pia, INFN Genova
Data libraries

- PIXE (proton ionisation cross sections)
- BEB – DM (electron ionisation cross sections)
- Photon elastic
  - RTAB (Lynn Kissel) trimmed and reformatted for MC use

R&D in progress

- Interpolation algorithms
- Smoothing techniques
- Data management methods

Experimentally validated
Quantified accuracy
State of the art

Mathematical expertise

Maria Grazia Pia, INFN Genova
Refactoring data management

- Today’s technology
  - …keeping an eye on the new C++ Standard

- Optimal container
- Pruning data
- Splitting files
- Software design
Experiment: Z. W. Bell (ORNL)

Validation

Statistical Decay

\[
\begin{array}{cccccccc}
  & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & \ln(2) & 20.0 \text{ min} & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & \ln(2) & 20.0 \text{ min} & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & \ln(2) & 1.3 \text{ min} & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & \ln(2) & 22.3 \text{ yr} & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & \ln(2) & 22.3 \text{ yr} & 0 \\
\end{array}
\]

New algorithm

Performance

\[
\begin{align*}
\text{Time for 10,000 chains [ms]}: & \\
\text{Classical:} & \uparrow \\
\text{Old:} & \uparrow \\
\text{Statistical:} & \downarrow \\
\text{End of chain – S:} & \uparrow \\
\text{End of chain – C:} & \uparrow \\
\end{align*}
\]

\[
\begin{align*}
\text{233U decay:} & \\
\text{233U, 229Th, 225Ra, 225AC, 221Fr, 217At, 231Bi:} & \\
\text{Geant4:} & \\
\text{refactored:} & \\
\text{new algorithm:} & \\
\end{align*}
\]

\[
\begin{align*}
\text{Counts:} & \\
\text{33Ba photo-peaks:} & 79.61 \text{ and } 80.99 \text{ keV} \\
\end{align*}
\]
Conclusions

A quantitatively validated, state-of-the-art Monte Carlo code is at reach.

Large effort
Supported by software design

Refactoring techniques and reengineering patterns contribute to improve **computational performance** and facilitate **validation**.

**Physics insight is the key**