Friedrich-Schiller-Universität Jena
Physikalisch-Astronomische Fakultät

Design and Implementation of Vectorized Pseudorandom Number Generators and their Application to Simulation in Physics

MASTER'S THESIS

for obtaining the academic degree

Master of Science (M.Sc.) in Physics

submitted by Markus Pawellek

born on May 7th, 1995 in Meiningen
Student Number: 144645

Primary Reviewer:	Bernd Brügmann
Primary Supervisor:	Joachim Gießen

Jena, July 13, 2019

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

List of Figures

Symbol Table

Symbol	Definition
$x \in A$	x ist ein Element der Menge A.
$A \subset B$	A ist eine Teilmenge von B.
$A \cap B$	$\{x \mid x \in A$ und $x \in B\}$ für Mengen A, B — Mengenschnitt
$A \cup B$	$\{x \mid x \in A$ oder $x \in B\}$ für Mengen A, B — Mengenvereinigung
$A \setminus B$	$\{x \in A \mid x \notin B\}$ für Mengen A, B — Differenzmenge
$A \times B$	$\{(x, y) \mid x \in A, y \in B\}$ für Mengen A und B — kartesisches Produkt
$\emptyset$	$\{\}$ — leere Menge
$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^n$	Menge der n-dimensionalen Vektoren
$\mathbb{R}^{n \times n}$	Menge der $n \times n$-Matrizen
$f \colon X \to Y$	f ist eine Funktion mit Definitionsbereich X und Wertebereich Y
$\partial\Omega$	Rand einer Teilmenge $\Omega \subset \mathbb{R}^n$
σ	Oberflächenmaß
λ	Lebesgue-Maß
$\int_\Omega f \,\mathrm{d}\lambda$	Lebesgue-Integral von f über der Menge Ω
$\int_{\partial\Omega} f \,\mathrm{d}\sigma$	Oberflächen-Integral von f über der Menge $\partial\Omega$
∂_i	Partielle Ableitung nach der i. Koordinate
∂_t	Partielle Ableitung nach der Zeitkoordinate
∂_i^2	Zweite partielle Ableitung nach i
∇	$\begin{pmatrix} \partial_1 & \partial_2 \end{pmatrix}^\mathrm{T}$ — Nabla-Operator
Δ	$\partial_1^2 + \partial_2^2$ — Laplace-Operator
$\mathrm{C}^k(\Omega)$	Menge der k-mal stetig differenzierbaren Funktion auf Ω
$\mathrm{L}^2(\Omega)$	Menge der quadrat-integrierbaren Funktionen auf Ω
$\mathrm{H}^1(\Omega)$	Sobolevraum
$f\|_{\partial\Omega}$	Einschränkung der Funktion f auf $\partial\Omega$
$\langle x, y\rangle$	Euklidisches Skalarprodukt
$[a, b]$	$\{x \in \mathbb{R} \mid a \le x \le b\}$
(a, b)	$\{x \in \mathbb{R} \mid a < x < b\}$
$[a, b)$	$\{x \in \mathbb{R} \mid a \le x < b\}$
$u(\cdot, t)$	Funktion $\tilde{u}$ mit $\tilde{u}(x) = u(x, t)$
A^T	Transponierte der Matrix A
id	Identitätsabbildung
$a \coloneqq b$	a wird durch b definiert
$f \circ g$	Komposition der Funktionen f und g
$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$	Determinante der angegeben Matrix
$\mathrm{span}\,\{\ldots\}$	Lineare Hülle der angegebenen Menge
$\|A\|$	Anzahl der Elemente in der Menge A

1 Introduction

2 Background

2.1 Stochastics and Statistics

2.2 Random and Pseudorandom Sequences

Pros and Cons about Pseudorandom numbers why do we not want pure random number generators? -> not deterministic how to get best of both worlds -> seeding

Define random sequences and pseudorandom sequences.

2.3 Random Number Generators

We should provide a rigorous definition of a random number generator. This will make things clear for the implementation and the new concept we are defining. Make sure to include seeding and parallel execution.

2.4 SIMD on Vector Processors

2.5 MIMD on Multiprocessors

3 Related Work

[11] [13] [9]

4 Methodology

4.1 Randomness Tests for RNGs

4.2 Performance of RNGs

There are 7 or 8 standard methods to parallelize RNGs. Explain the methods and their pros and cons. Say something about if it is better to use them vectorized or in multiprocessor.

4.2.1 Vectorization

4.2.2 Parallelization

4.2.3 Aliasing, Caching and Memory

4.3 Assembly Language

4.4 High-level Language

4.5 Intel Intrinsics

Naming scheme direct map to assembler instructions references: intel intrinsic guide advantages and disadvantages library abstraction through xsimd or agner fogs vector library

4.6 Libraries

4.7 C++ Concepts for Random Number Generators

Advantages and Disadvantages Better ideas taking best of all worlds

5 Implementation

5.1 Project Structure

5.2 Tools and Libraries

godbolt google benchmark intel vtune amplifier testu01 dieharder practrand cache miss measure by agner fog different compilers

5.3 Interface

What do we want from the interface of our RNG? It should make testing with given frameworks like TestU01, dieharder, ent and PractRand easy. Benchmarking should be possible as well. Therefore we need a good API and a good application interface. Most of the time we want to generate uniform distributed real or integer numbers. We need two helper functions. So we see that the concept of a distribution makes things complicated. We cannot specialize distributions for certain RNGs. We cannot use lambda expressions as distributions. Therefore we want to use only helper functions as distributions and not member functions. So we do not have to specify a specialization and instead use the given standard but we are able to do it. Therefore functors and old-distributions are distributions as well and hence we are compatible to the standard.

Additionally, we have to be more specific about the concept of a random number engine. The output of a random number engine of the current concept is magical unsigned integer which should be uniformly distributed in the interval [min,max]. But these magic numbers can result in certain problems if used the wrong way, see Melissa O'Neill Seeding Surprises. Therefore the general idea is to always use the helper functions as new distributions which define min and max explicitly and make sure you really get those values. This is also a good idea for the standard. And it is compatible with the current standard.

Now think of vector registers and multiprocessors. The random number engine should provide ways to fill a range with random numbers such that it can perform generation more efficiently. Think about the execution policies in C++17. They should be provided as well.

6 Results

godbolt google benchmark intel vtune amplifier testu01 dieharder

7 Conclusion

References

1. Barash, L. Y. & Shchur, L. N. PRAND: GPU accelerated parallel random number generation library: Using most reliable algorithms and applying parallelism of modern GPUs and CPUs. *Computer physics communications* **185,** 1343–1353 (2014).
2. Barash, L. Y. & Shchur, L. N. RNGSSELIB: Program library for random number generation. More generators, parallel streams of random numbers and Fortran compatibility. *Computer Physics Communications* **184,** 2367–2369 (2013).
3. Barash, L. Y. & Shchur, L. N. RNGSSELIB: Program library for random number generation, SSE2 realization. *Computer Physics Communications* **182,** 1518–1527 (2011).
4. Barash, L. Y., Guskova, M. S. & Shchur, L. N. Employing AVX vectorization to improve the performance of random number generators. *Programming and Computer Software* **43,** 145–160 (2017).
5. Bauke, H. & Mertens, S. Random numbers for large-scale distributed Monte Carlo simulations. *Physical Review E* **75,** 066701 (2007).
6. Demirag, Y. *Vectorization Studies of Random Number Generators on Intel's Haswell Architecture* (CERN openlab, 2014).
7. Fog, A. *Calling Conventions for Different C++ Compilers and Operating Systems* https://www.agner.org/optimize/calling_conventions.pdf (Agner Fog, 2018).
8. Fog, A. *Optimizing Software in C++: An Optimization Guide for Windows, Linux and Mac Platforms* https://www.agner.org/optimize/optimizing_cpp.pdf (Agner Fog, 2018).
9. Fog, A. Pseudo-Random Number Generators for Vector Processors and Multicore Processors. *Journal of Modern Applied Statistical Methods* **14.** http://digitalcommons.wayne.edu/jmasm/vol14/iss1/13 (2015).
10. Guskova, M. S., Barash, L. Y. & Shchur, L. N. RNGAVXLIB: Program library for random number generation, AVX realization. *Computer Physics Communications* **200,** 402–405 (2016).
11. Kneusel, R. T. *Random Numbers and Computers* ISBN: 978-3-319-776696-5 (Springer, 2018).
12. L'Ecuyer, P. *Random number generation with multiple streams for sequential and parallel computing* in *2015 Winter Simulation Conference (WSC)* (2015), 31–44.
13. Saito, M. & Matsumoto, M. in *Monte Carlo and Quasi-Monte Carlo Methods 2006* 607–622 (Springer, 2008).

A Build System

[8]

Statutory Declaration

I declare that I have developed and written the enclosed Master's thesis completely by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. The Master's thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.
On the part of the author, there are no objections to the provision of this Master's thesis for public use.

Jena, July 13, 2019

Markus Pawellek