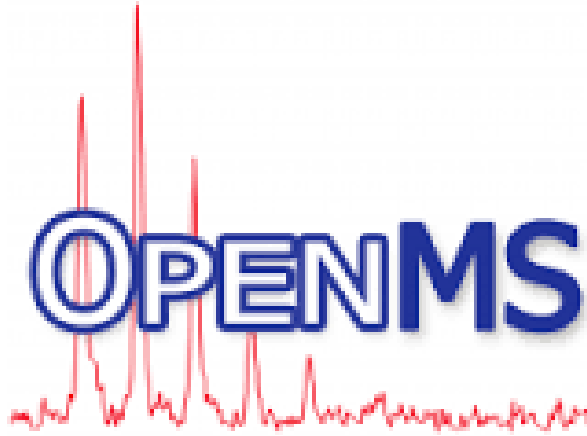


Projektmanagment im Softwarebereich



Performance Tuning of Algorithms
in
High-Throughput Proteomics

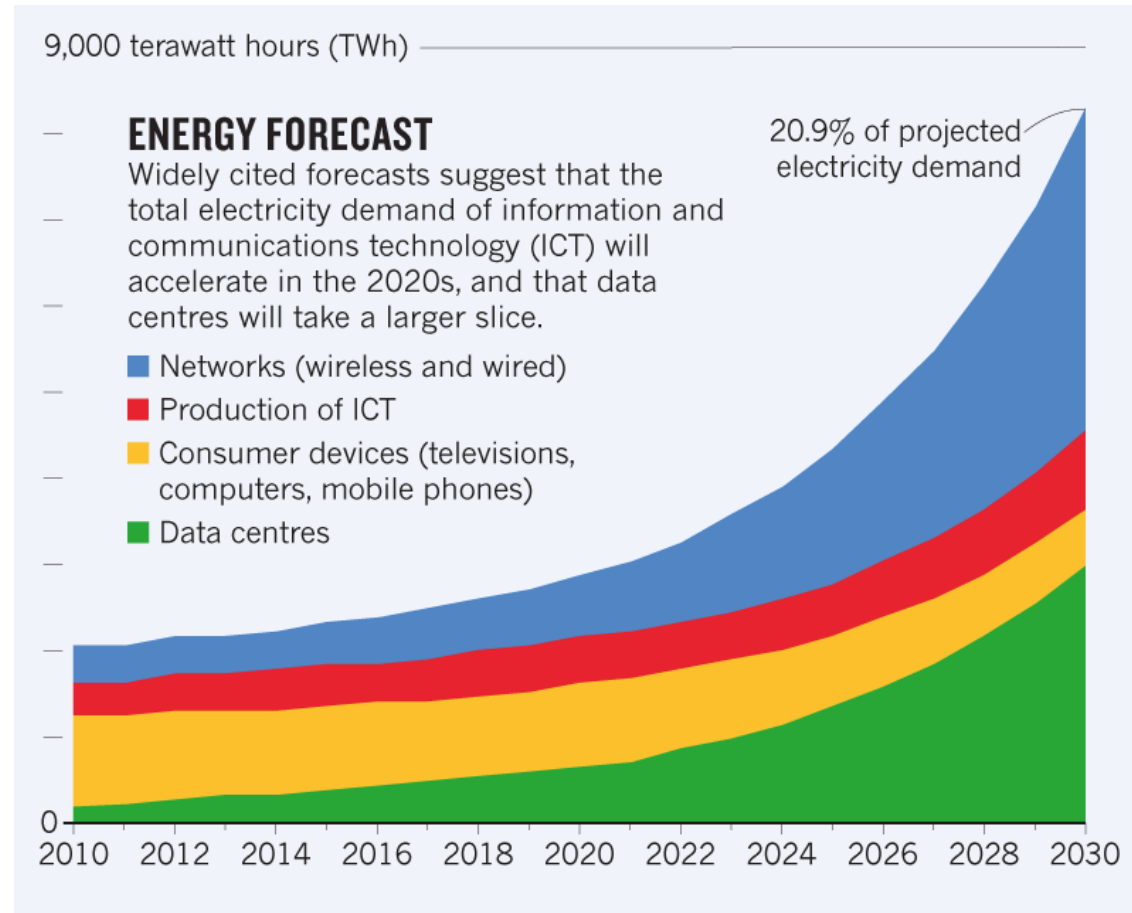
Chris Bielow

Warum Performance?

- Gut für den Akku
 - Laptops, Smartphones, ...

Warum Performance?

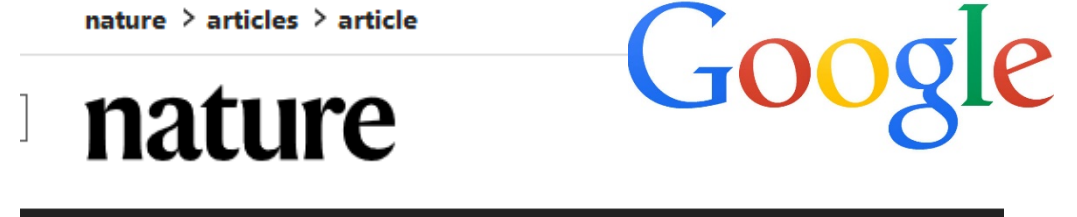
- Gut für den Akku
 - Laptops, Smartphones, ...
- Gut für unser Klima
 - ...



How to stop data centres from gobbling up the world's electricity, Nature **561**, 163-166 (2018)
doi: 10.1038/d41586-018-06610-y

Warum Performance?

- Gut für den Akku
 - Laptops, Smartphones, ...
- Gut für unser Klima
 - ...
- Gut für's Image
 - Quantum Supremacy
- Gut für den Benutzer
 - High-Performance-Computing



IBM

IBM Research Blog

*We argue that an ideal simulation of the same task can be performed on a classical system in **2.5 days** and with far greater fidelity.*

cy using a
perconducting

artinis

article

660k Accesses | 26 Citations | 6028 Altmetric | Metrics

“a state-of-the-art supercomputer would require approximately **10,000 years** to perform the equivalent task.”

Proteomics

Genomics

Transcriptomics

“OMICS”

Connectomics

Foodomics

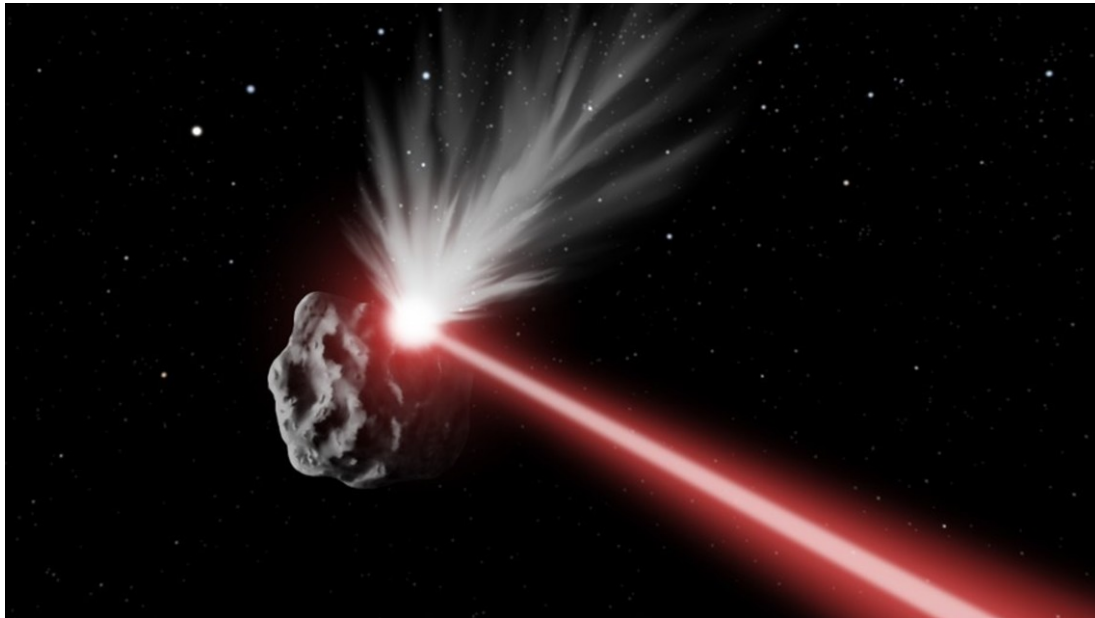
Metabolomics

OMICS – Ganzheitliche Charakterisierung und Quantifizierung der Moleküle einer Zelle/Probe/Organismus

Proteomics

Method of choice:

Liquid Chromatography – Mass Spectrometry (LC-MS)

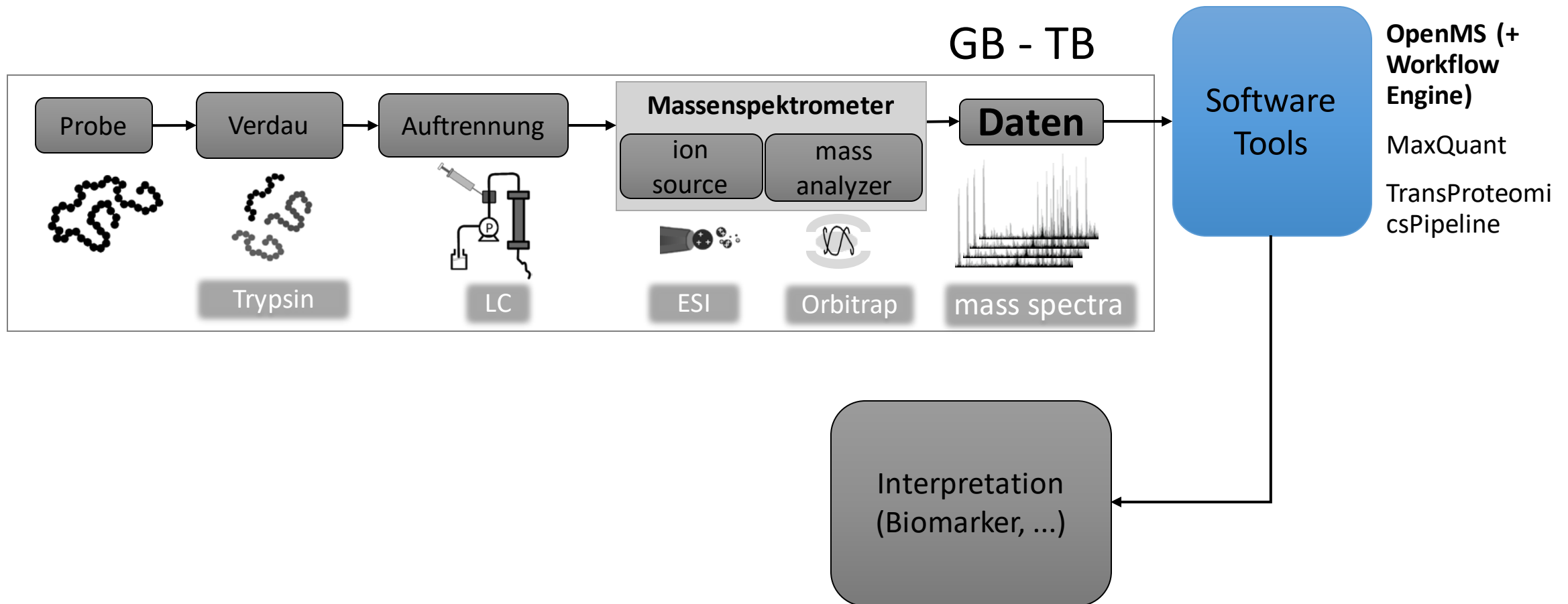


“Shooting at stuff
with lasers in space”!

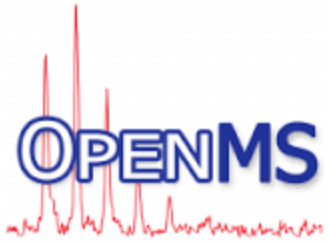
https://www.syfy.com/sites/syfy/files/styles/1140x640/public/2017/06/170428-debris-laser-mn-1605_c382498d2eef0ff260fc6c8f70ca4806.focal-920x600.jpg

Proteomics-Pipeline

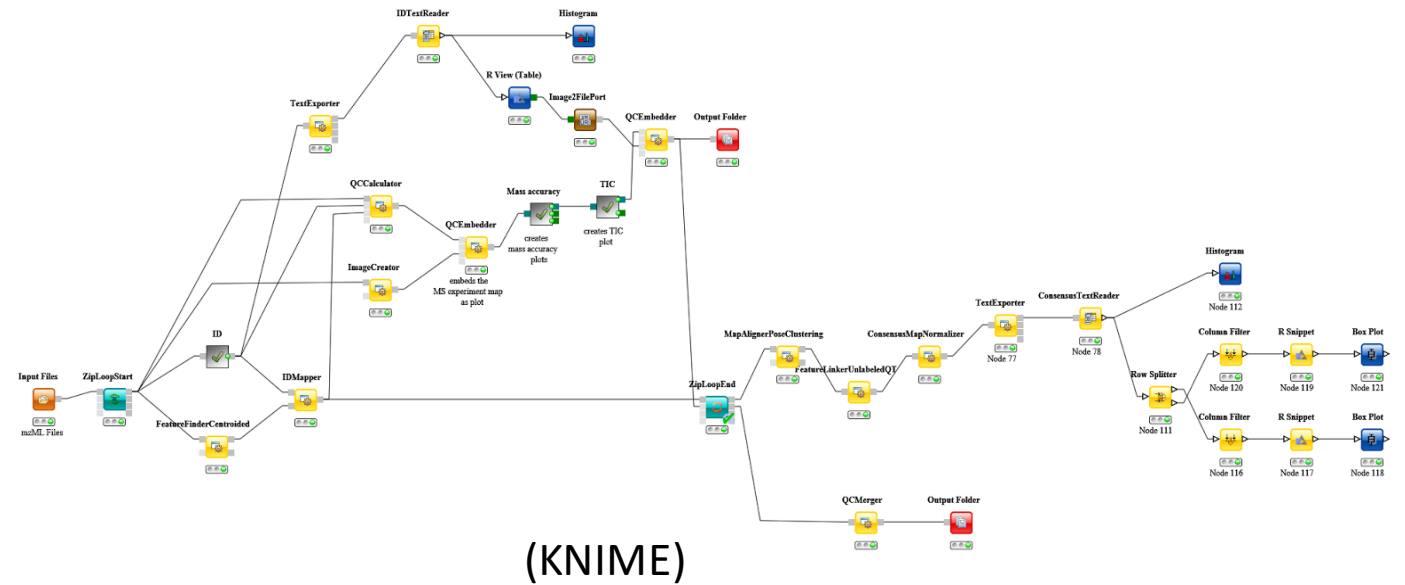
Liquid Chromatography (LC) + Massenspektrometrie (MS)




Software Pipeline



- C++ Bibliothek für Analyse von LC-MS Daten in der Proteomik
- Tools/Executables
 - 150 Tools
 - „Lego“



Inhalt des Praktikums

- Schrittweises Erlernen von Techniken zur Verbesserung der Laufzeit
 - Do/Don't Idioms in C++ (Wahl von Algorithmen und Datentypen, Datenstruktur, Allokatoren, Dispatching)
- Optimierung der Programmlaufzeit
 - C++ Code Profiling: Intel V-Tune, perf,...
 - CPU architecture: cache und SIMD/AVX instructions
- Benchmark-Umgebungen
 - Microbenchmarks (Google benchmark)
- Anwendung auf Tools in der Proteomik-Pipeline von OpenMS (im Team)
- Arbeit in bestehender open-source Bibliothek (BSD-Lizenz)
 - Erstellung von Pull-Requests ( **GitHub**) und Code Reviews
 - Aufnahme der neuen Funktionalität in die OpenMS-Entwicklerversion
- Echte Daten, echte Probleme, echte Ergebnisse
- Algorithmen in der Proteomik

Zeitplan - Seminar

Ende Februar ...

- Vorbesprechung, genaue Terminvereinbarung per E-Mail zur Vermeidung von Terminkollisionen

Mo, 09.03. - Fr, 13.03. – Projekteinführungswoche

- Projektmanagement, Zeit & Ressourcen-Planung
- Einführung in Flüssigkeitschromatographie & Massenspektrometrie (LC-MS)
- Refresher: C++ && Arbeit in einer großen Softwarebibliothek
- Tutorial OpenMS (C++ library und Tools)
- Workflow engines: TOPPAS
- Git (version control software)
- Einführung in 'perf' & Co; flame graphs, ...

Zeitplan – praktischer Teil

Fr, 22.03.

Vorstellung der Pläne

Mo, 23.03. - Fr, 01.05.

Woche 1(+2):

- OpenMS-develop installieren
- Projekt auf <https://git.imp.fu-berlin.de> clonen
- Einarbeitung in OpenMS
- Relevante Literatur lesen
- (Externe Programme testen)
- Gemeinsam Zielalgorithmen aussuchen, Zeitplan erstellen, Vortrag

Woche 3+4

- Tuning/Profiling
- Implementierung (C++)
- Continuous Integration...

Woche 5

- Code finalisieren (C++); evtl. Bugs fixen; finaler PR an OpenMS
- Abschlussbericht schreiben (Erreichte Verbesserungen, Plots)

Fr, 08.05. Abschlussbericht und Vorstellung der Ergebnisse

wöchentliche Besprechung

Freitags: 10-12 Uhr,

Weiteres Vorgehen, Anpassung des Zeitplans, Problemlösung

Weitere Treffen nach Bedarf

Vorraussetzungen

- Erfahrung in Objektorientierter Programmierung (Java, C++, ...)
 - C++ Kenntnisse (wenig Templates)
 - z.B. AIDaBi Praktikum / C++-Kurs (Nachmeldung mgl., via Prüfungsbüro Formular)

Empfehlenswert

- Eigener Laptop
 - Zum Entwickeln (50 GB frei, 4C)
 - Als thin-client
 - Zugang zu Compute-Servern 28C wird gewährt)

Schwierigkeitsgrad (Acht Sterne verteilt auf drei Bereiche):

A Programmieren: *****

B Biologie/Chemie: *

C Projektmanagement: **

Fragen?

- mehr Fragen
 - chris.bielow@fu-berlin.de

Ausbau zu B.Sc. Arbeit
prinzipiell möglich

Mehr Informationen

- OpenMS
 - www.OpenMS.de
- PMSB OpenMS
 - [https://www.bsc.fu-berlin.de/TeachingAndWorkshops/SoSe20/Software Praktikum OpenMS/index.html](https://www.bsc.fu-berlin.de/TeachingAndWorkshops/SoSe20/Software_Praktikum_OpenMS/index.html)

BTW:

Google Summer
of Code

