ACAT 2021
Summary

Axel Naumann, CERN axel@cern.ch
2021-12-03
ACAT 2021
29 NOV - 03 DEC, 2021

AI Decoded:
Towards Sustainable, Diverse, Performant and Effective Scientific Computing

Daejeon, Republic of Korea
Virtual and IBS Science Culture Center
Numbers: Abstracts

- 245 abstracts - likely a record high
- Most abstracts from Europe: something we will want to improve in the future!

**Europe** 62%

- Asia 11%
- America 11%
- No preference 17%
Numbers: Contributions

- 14 plenary
- 136 posters
- Full parallel tracks: 4 days times 6 talks = 24 T3, 25 T1, 26 T2 ?!
Challenge

• *Zoom fatigue (n)*

• *Wikipedia entry, thus a thing*
Challenge

• Time zones
## Time zones

<table>
<thead>
<tr>
<th>Time</th>
<th>Daejeon</th>
<th>Moscow</th>
<th>Berlin</th>
<th>New York</th>
<th>Seattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-9:30</td>
<td>0-3:30</td>
<td>22-1:30</td>
<td>16-19:30</td>
<td>13-16:30</td>
<td></td>
</tr>
<tr>
<td>9-12:30</td>
<td>3-6:30</td>
<td>1-4:30</td>
<td>19-22:30</td>
<td>16-19:30</td>
<td></td>
</tr>
<tr>
<td>12-15:30</td>
<td>6-9:30</td>
<td>4-7:30</td>
<td>22-1:30</td>
<td>19-22:30</td>
<td></td>
</tr>
<tr>
<td>15-18:30</td>
<td>9-12:30</td>
<td>7-10:30</td>
<td>1-4:30</td>
<td>22-1:30</td>
<td></td>
</tr>
<tr>
<td>18-21:30</td>
<td>12-15:30</td>
<td>10-13:30</td>
<td>4-7:30</td>
<td>1-4:30</td>
<td></td>
</tr>
<tr>
<td>21-0:30</td>
<td>15-18:30</td>
<td>13-16:30</td>
<td>7-10:30</td>
<td>4-7:30</td>
<td></td>
</tr>
</tbody>
</table>
Solution

- Short sessions, much condensed program
- ACAT is a workshop, want discussions!
  20 minutes per slot, significantly less presentations
- Most (and excellent!) content in poster session
- Many abstracts: high productivity yet few occasions to present
  - Painfully small oral acceptance rate...
Participants

- 603 registered participants
- Complex time table
- Hybrid workshop
- Actual attendance around 175 in Asia/Europe sessions, 75 in Asia/Americas sessions
대전시 and the Conference Area
Zooming into the Local Venue
Right before we started
Agenda, Data, Videos

- Massive help from CERN

- Yasemin Altinbilek instrumental for arranging Indico, timetable, plenaries, registrations, etc: **THANK YOU!**

- Hybrid event with support from CERN - what does that mean for data protection? Complexity! Crucial support from CERN's Office of Data Privacy: **THANK YOU!**

- Serving days worths of recordings with CERN's video service: **THANK YOU!**
Poster session

- Many busy poster stands
- Low ratio of audience / presenter: would need more sessions
- Very high poster quality
- Enjoyed excellent presentations, interaction, and discussions: Gather.Town works!
Best poster and interaction

- Congratulations again to the best posters!
- Congratulations also to "best participants":
  - Felice Pantaleo
  - Ioana Ifrim
  - Ivan Kabadzhov
What is "Local"?

• "Local" organizers did an incredible job: smooth and elegantly

• "We have learned how to do conferences during COVID"

• Maybe, but completely relies on a fantastic LOC like at ACAT 2021!

From the whole SPC:

THANK YOU, LOC!
Tracks' Summary
Track 1: Computing Technology for Physics Research

• A big **THANK YOU** to those who organized Track 1:
  
  • Chang-Seong Moon, Kyungpook National University
  • Gordon T. Watts, University of Washington
  • Marilena Bandieramonte, University of Pittsburgh
  • Patricia Mendez, CERN
Track 2: Data Analysis - Algorithms and Tools

- A big **THANK YOU** to those who organized Track 2:
  - Erica Brondolin, CERN
  - Jennifer Ngadiuba, Fermilab
  - Junghwan Goh, Kyung Hee University
  - Liliana Teodorescu, Brunel University London
Track 3: Computations in Theoretical Physics: Techniques and Methods

- A big **THANK YOU** to those who organized Track 3:
  - Andrey Pikelner, JINR
  - Daniel Maitre, Durham University
  - Maria Ubiali, University of Cambridge
  - Myeonghun Park, Seoultech
Summary's Goal

• Based in tracks' input:
  • Summarize the main themes of the tracks
  • Determine future trends
  • Combine and synthesize
• I might have misunderstood the experts or the crystal ball
• Apart from possible errors, the following predictions are guaranteed!
Contributions

• Calling out some contributions, not a function of quality of contribution: apologies!

• All contributions (plenary and parallel) appearing as recordings:

  **THANK YOU**, Jonas Hahnfeld, Javier Lopez Gomez, Vincenzo Padulano!
Track 1
Track 1 Highlights

- Ongoing innovation through ML applied to computing: observation, tuning
- Optical processing units! #634
- Trigger with GPU, without ML! (LHCb, CMS, ATLAS)
- Z-vertex ML finding (Belle II) #750
- FPGAs for MC simulation! (QED, simpler geometry) #731
- GPU? MPI? tbb? JUNO analysis: all of them! #609
- Compile-time polymorphic detector description: for tracking, also for simulation? #739

From #739: detray - A compile-time polymorphic tracking geometry description
Track 1 Themes and Challenges

• I/O remains to be optimized: request/response buffering system; caching #645 #725 #760; distRDF; compression #678; RNTuple #659; data carousel #542; merging is part of this and remains an issue for concurrent processing #578

• Efforts to move from ROOT TTTree to numpy arrays; performance remains a danger (GIL, data processing in python); interoperability through C-arrays #587, #618, #765

• Hardware vs software evolution: continuously adjusting our code to newest generations is not free, despite abstractions
Track 1 Outlook

• Staying heterogeneous, remains a (really just engineering?) challenge: GPU! FPGA! Maybe ASIC! Move data around ("I/O"?); scheduling: keep everyone busy

• ML applied to computing, data layout, and I/O

• Columnar data (used by HENP since decades!) is in focus for analysis. Columnar data formats not a topic anymore - but analysis ergonomics?

• Dask and Spark and whatever else: scheduler abstractions, benchmarking, applicability for our workloads. Challenges remain scheduling and slicing.
Track 2
Track 2 Highlights

• Almost exclusively ML abstracts these days
  • Diverse topics within ML

• Impossible to select reasonable number of highlights

• Tracking for LUXE@DESY #580: comparing all existing algorithms classes, classical ACTS, ML, QC

• Fitting and performance (incl. automatic differentiation) #666 #741 and track 1's #599
Track 2 Themes

• GPU: many discussions, people seem to have similar / overlapping problems. ACAT wonderful venue for that (even better in person...)

• ML well beyond that stage: use with aplomb

• QC not there yet
Track 2 Future Challenges

• Entrance ticket into even discussing QC is expensive: it's a real investment to even understand and be able to engage in discussions
  • Is this different from ML? (Likely, due to the fundamental shift in algorithms)

• ML on a next level: ingredients understood, people are architects creating sophisticated buildings out of these ingredients
  • Auto encoders will remain a hot topic, also for anomaly search

• ML on real data, not just MC! HL-LHC effect, or deficit of ML vs production?
Track 2 Future Challenges

- ML in production
  - changes in detectors, efficiency etc?
  - More focus on designing efficient, fast algorithms, less on robustness?
  - Development focus exclusively on ML, alternatives are neglected because less sexy, yet we understand the robustness of traditional ones. See e.g. Gordon's comment on floating point precision versus disturbing ML results
Track 3
Track 3 Highlights

- Quantum circuits for HEP computations: early but promising #532
- ML + Bayesian: matrix element calculation #768, MC integration #648
- Normalizing flows to extract probability density from data #657
- One of a kind: lattice QCD, checking presence of renormalons in QCD perturbative expansions #582
- Important plenaries such as Anja Butter and Kevin Buzzard: understanding boundary between human computation, computers, and AI is useful!
Track 3 Theme

• Codes becoming open source! #778 #615
• Even for complex code like PDF fitting
• No sausage factory ("don't know what's inside")!
• Milestone for reproducibility!
• NNNLO calculations require advances also in PDFs
Track 3 (Future) Challenges

- QC for HEP problems: e.g. for matrix elements for MC event generation? From QCD Lagrangian to event generation?
  - QC applications before we have the machines! Yippee? Scaling requires actual QC machines!
- NNLO automation; ML will become common, level out at some point
- Efficiency! Wide spectrum of tools; focus on throughput (negative weights etc)
- Tweaking simulation to match data: Geant4, event generators / PDF fits (Butter): using theory vs or even generation from data? Interpretation?
Track 3 at ACAT

- ML, computing challenges, accelerators: track 3 is well connected with track 1 and track 2!
- Very happy to see consistent participation from generator community!
Summary of ACAT 2021
Common Challenges

- GPUs vs our code and algorithms (and problems?), e.g. simulation; beneficial for some elements (trigger, automatic differentiation, of course ML)

- HPC to production

- Software sophistication vs efficiency

- None of this is new, but takes decades (of hard work!) from discussions + prototypes to production quality

By Stefano Bertolotti - Own work, CC BY-NC-SA 2.0, https://www.flickr.com/photos/cicciopizzettaro/4297936934
Life of a Post-Doc

- Ruth Müller #786: *Racing for what? Acceleration and Competition in the Academic Sciences*

- Post-doc: "I was surprised that people didn't seem to be aware?"

- We have a lot of work ahead of us, on many areas (no, not just technical)!

- Isn't that wonderful?

ACAT 2021
29 NOV - 03 DEC, 2021
AI Decoded:
Towards Sustainable, Diverse, Performant and Effective Scientific Computing
Daejeon, Republic of Korea
Virtual and IBS Science Culture Center
Algorithms at the core, communicating arcs connecting the foundations:
at ACAT, you build the bridge for HENP