Agile Research - Strengthening Reproducibility in Collaborative Data Analysis Projects

Till Moritz Karbach[1], Christian Peter Linn[1], Sebastian Neubert[1], Andrey Ustyuzhanin[2,3,4]

[1] CERN, [2] Yandex School of Data Analysis, [3] HSE, [4] NRC "Kurchatov Institute"

CHEP 13.04.2015

守破離

Why?

My Analysis

What do I measure?
Which effects to take into account?
How do I build the analysis?

Me

Professor
(writing grant apps)

What do I measure?
Which effects to take into account?
How do I build the analysis?

Same Analysis

My Analysis

What do I measure?
Which effects to take into account?
How do I build the analysis?

Me

WG Convenor

Professor
(Writing grant apps)

Our Analysis

What do we measure?
Which effects to take into account?
How do we build the analysis?

Team

What Do We Need to Improve?

- Reproducibility of results
- Tackling more complex analyses
- Sustainability and Happiness
- Communication soft skills

"The skills required to be a successful scientific researcher are increasingly indistinguishable from the skills required to be successful in industry"

Jake Vanderplas, 2013, `http://bit.ly/19FmPex`

Pilot Project: A Scrum Analysis Team

How to apply agile concepts in the real HEP world

"In computational research it will be nothing less acceptable as a result than data and executable code"

David L. Donoho et al, 2009, `http://stanford.io/1cyyfSd`

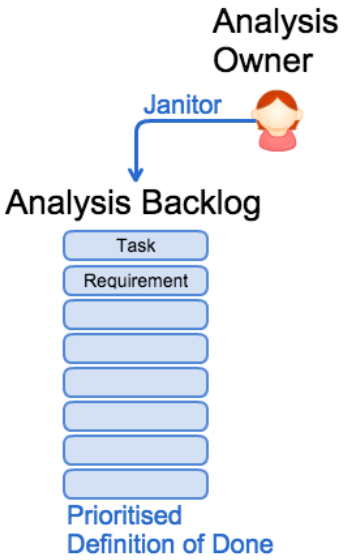
Scrum Analysis Team

Pilot project

- 3 Summerstudents + 3 CERN Fellows
- 2 Analysis topics: Search for new decays of
  - $\Lambda_b \to \Lambda_c^+ K^+ K^- \pi^-$,
  - $B_s \to f_2(1525)\mu\mu$
- Timeframe: 13 weeks
- 2-day Kick-off workshop
- Testing Scrum as a development framework

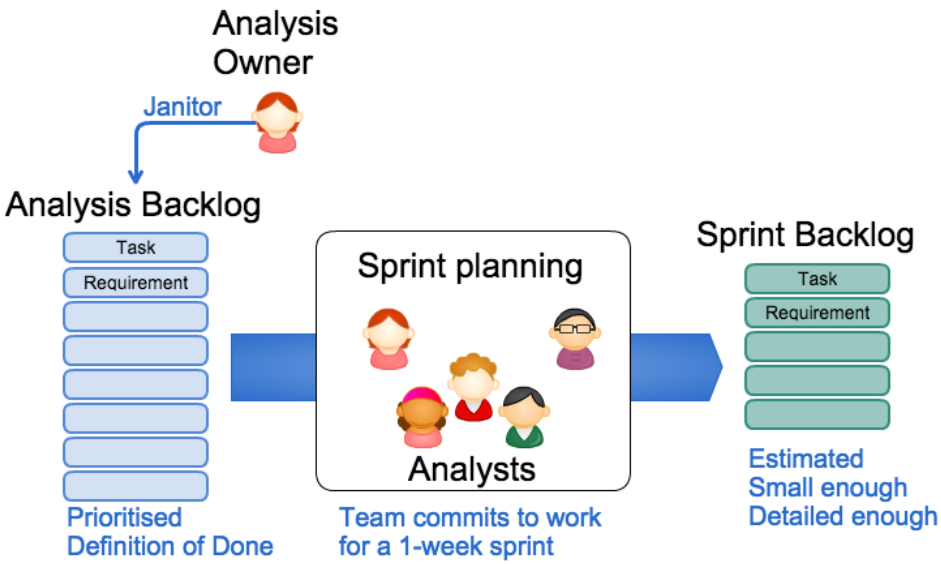
What is SCRUM?

A Framework for Continuous Improvement



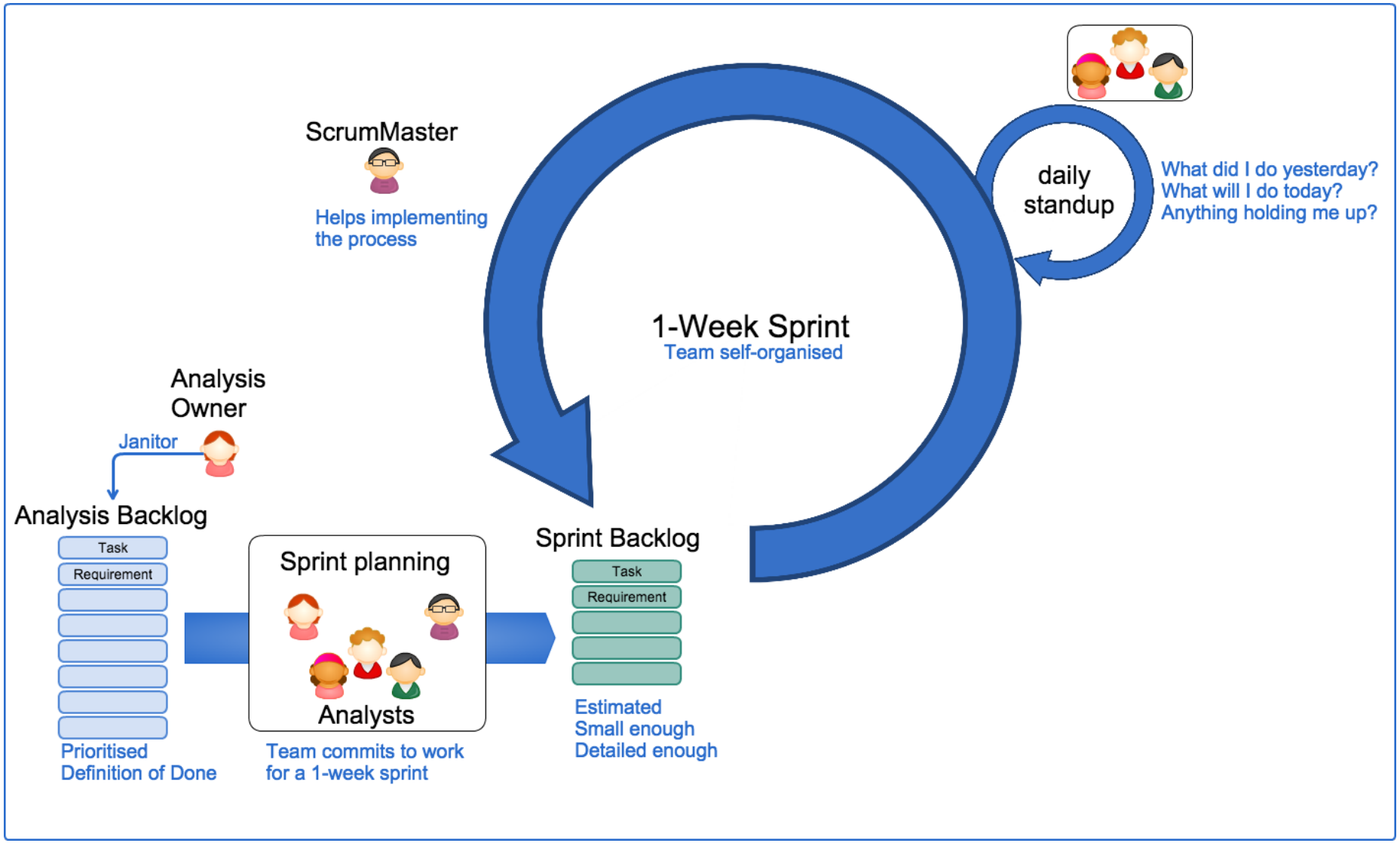
What is SCRUM?

A Framework for Continuous Improvement



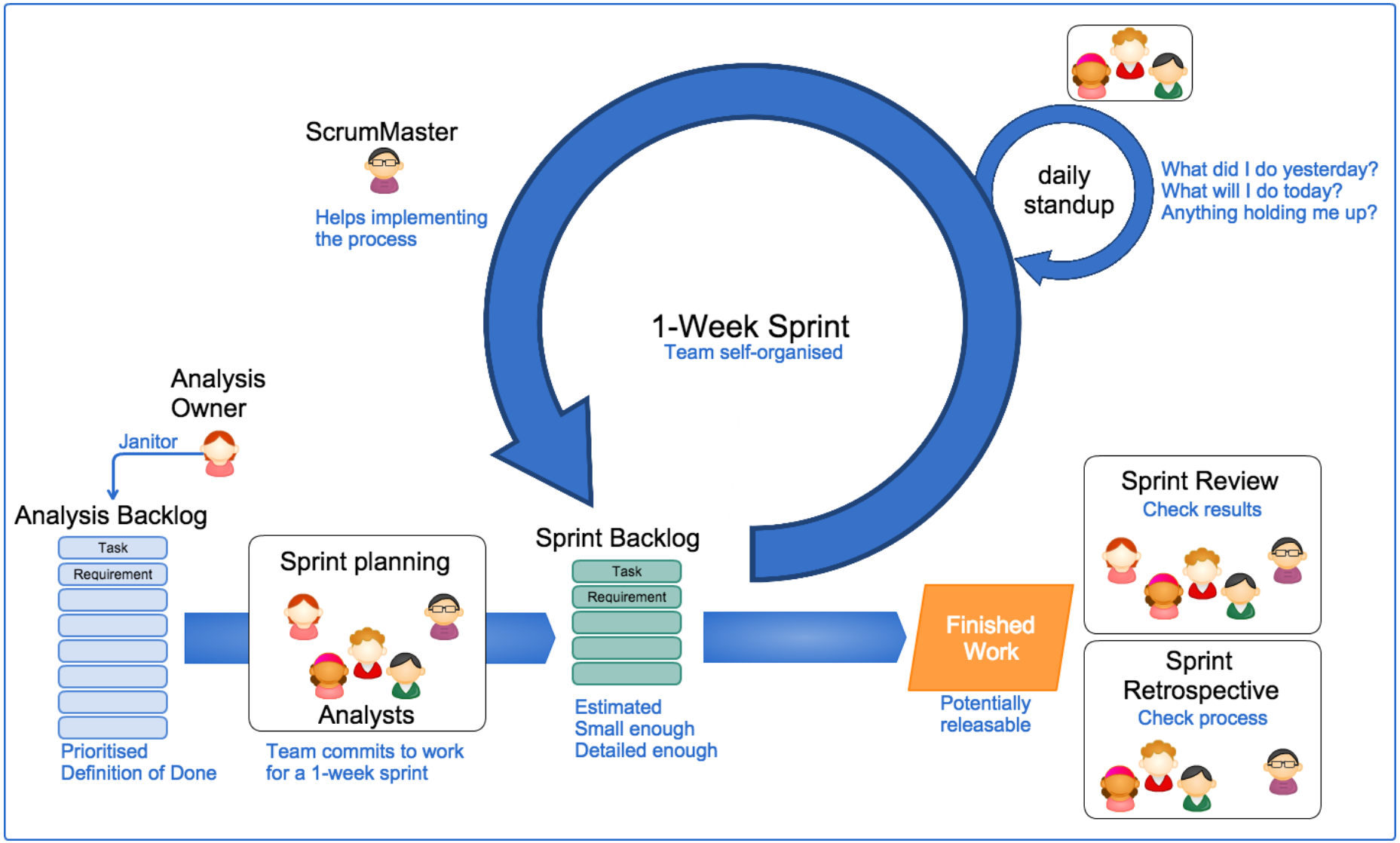
What is SCRUM?

A Framework for Continuous Improvement



What is SCRUM?

A Framework for Continuous Improvement



A Virtual Pinboard

Visual organisation of tasks with ↪ Trello

SummieProject: First Obse ×

https://trello.com/b/MstQoFac/summieproject-first-observations-at-lhcb

Apps | https://mail.google | Trello | Indico [LHCb Meetin | Shift/Piquet | LHCb Coding | Private | Cacoo – Diagrams | Git Cheatsheet • ND | Diff Checker – Onlin | Other Bookmarks

Boards | Trello | SN Sebastian Neubert

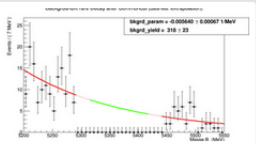
SummieProject: First Observations at LHCb ☆ Private | Burndown Chart | Subscribed 5 | Show sidebar

Backlog

- compare kinematic ranges of signal and PID calib samples (protons!)
- Lb: Nominal fit to extract the signal yields.
- check data / MC agreement - can we use MC for the PID kinematic reweighting?
- PID calib: identify a sensible binning for the reweighting of the kinematic variables: signal vs PID calib; data/MC; signal/control sample
- compute PID efficiency of PID cuts using PID calib samples
- Compute branching fraction.
- Fit validation?
- Limit setting?
- Systematics
- Should we use the MC to train the BDT?
- Kk system non resonant component

Add a card…

Sprint Backlog

- Bs: Get kinematic efficiency from MC for NON-RESONANT Decay: geometric, trigger, selection — 6 | 1 | 0/1 — DB
- 
  Bs: From efficiencies calculate expected non-res. Signalevents — 2 | 10 — DB
- fitter2: Add function to define ranges (fitrange, signalrange)
- fitter2: Allow 2D-fits
- Lb: Test if the BDT could be improved by using sidebands rather than sweights. — VA

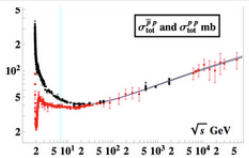
Add a card…

In progress

- Look at Lb CP asymmetry. — 5/6 — VA
- Documentation — 1 | 2 | 1/2 — VA DB

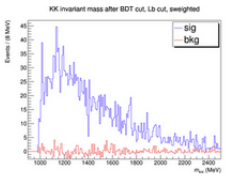
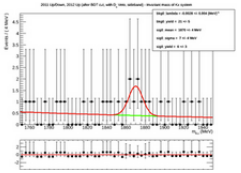
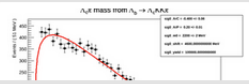
Add a card…

For Sprint review

- Reproducing 2012 ntuples Lb2LcKKpi on correct stripping — 3 | 4/4 — SN
- 
  Understand p-pbar detection asymmetry — 1 | 1 — VA
- Reproduce Lb decays data together with 2012MagDown — VA
- Lb two body: tabelle yields — VA
- Lb two body: 2D plots — VA
- Lb two body: calculate efficiencies (common FOM) — VA
- Combine Lb 2012MagDown data into one file — 2 — VA

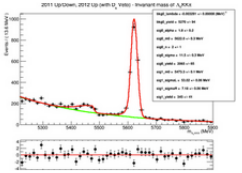
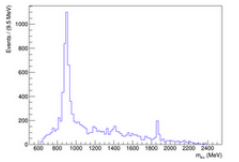
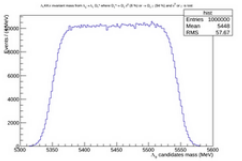
Add a card…

Done by 12.09.14

- summary two body subsystems analysis results — 1 — VA
- Add Mirrored Argus with shift into fitter — VA
- 
  Lb: 2 body plots with sweights (before and after BDT) — 7 — VA
- 
  Prove two body subsystems come from Lb: sideband plots — 8 — VA
- 

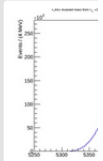
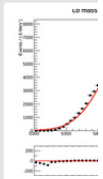
Add a card…

Done by 05.09.2014

- Presentation: summary Lb fit results — 1 — VA
- 
  Optimize the cut on the bdt response using s/sqrt(s+b) — 2 | 6 — VA
- 
  2 body masses after bdt cut — 9 — VA
- 
  Simulation Lb->Ds* L_c with one

Add a card…

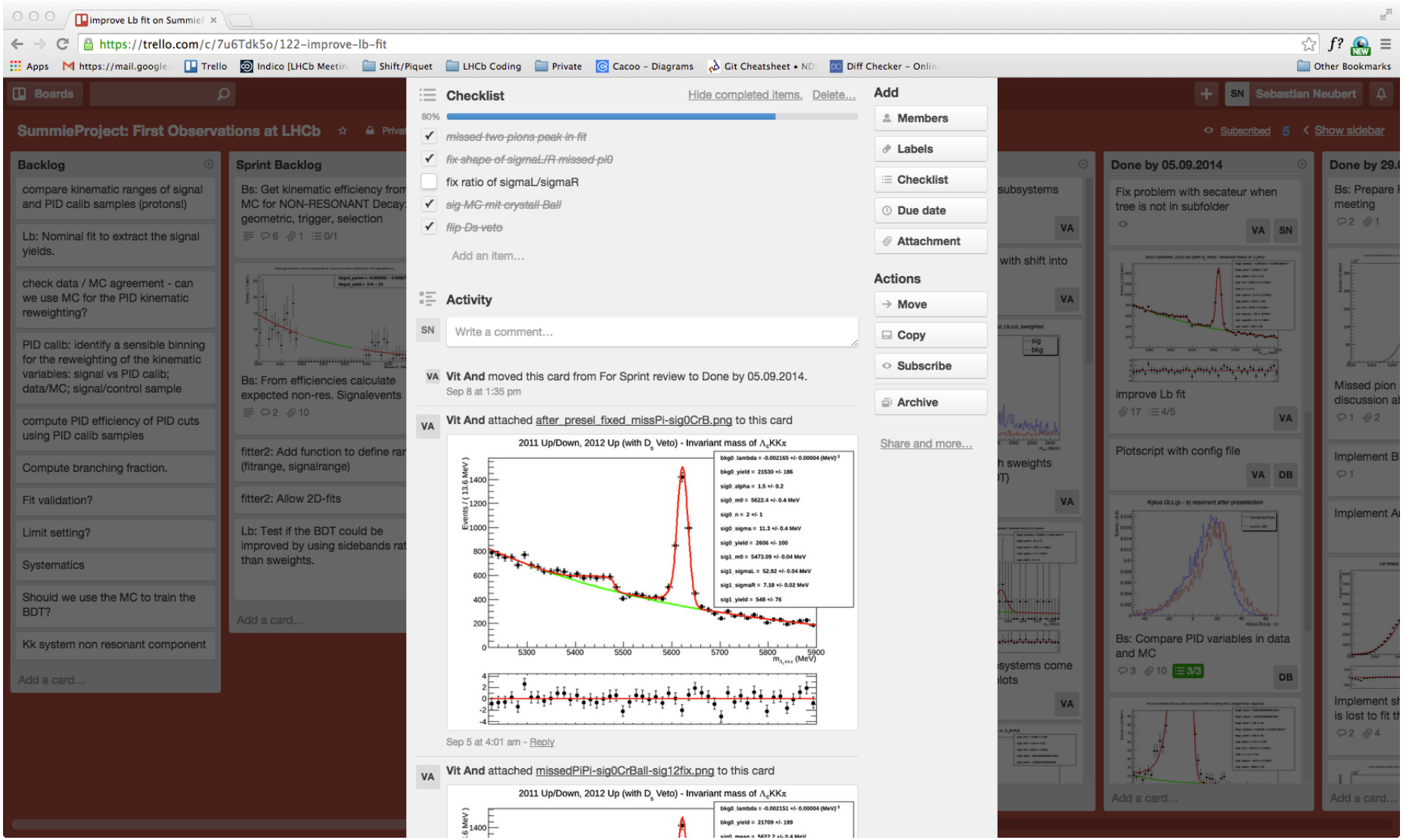
Done by 29.

- Bs: Prepare … meeting — 2 | 1
- 
  Missed pion … discussion al — 1 | 2
- Implement B — 1
- Implement A
- 
  Implement sl… is lost to fit th — 2 | 4

Add a card…

A Virtual Pinboard

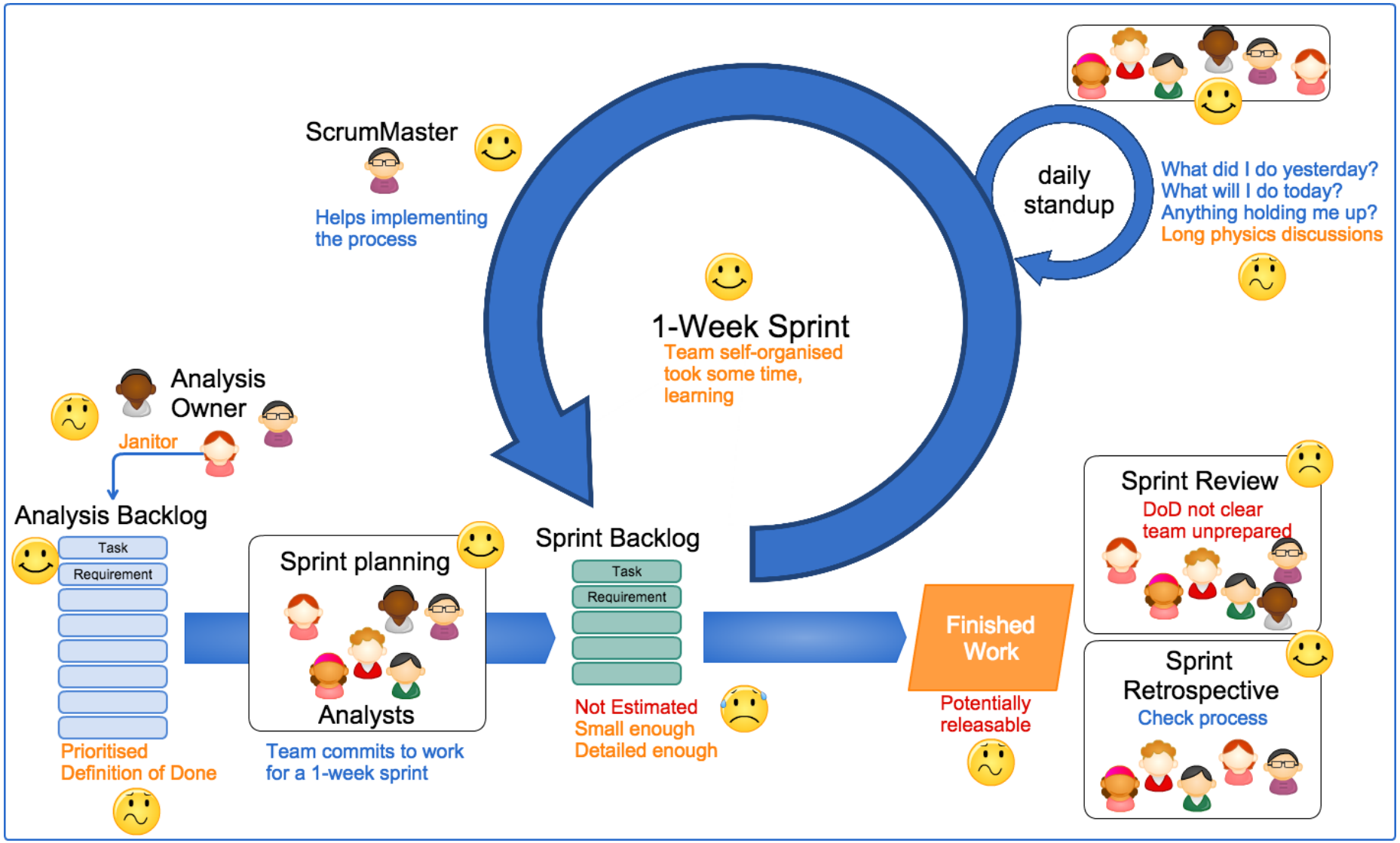
Visual organisation of tasks with ↪ Trello



Project Evaluation

How well did we implement SCRUM?

Summary from Survey Part I



Things we could have done better

Impacts of Implementation Failures

Poor task decomposition

- Team had to ask a lot of clarifying questions during sprint
- Often additional tasks created ad hoc
- $\Rightarrow$ insecurity how to organise learning

No estimates

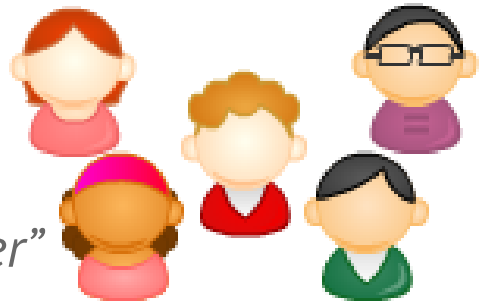
- Missed learning opportunity
- Little idea what to expect in one sprint
- No idea how much time needed for learning

Definition of done not clearly defined/checked

- Never really sure when things were finished
- Physics discussions slowed down by incomplete results

Impact: Teamwork

What helped to work as a team? Where did teamwork help the project?



- Daily meetings helped, very important
- Team *"needed a couple of weeks to get used to each other"*
- Diverse team with different ideas what to get out of the project
- Tools could be shared and reused
  - *"Explaining my tools to other team members helped me develop better tools"*
  - Enabled finish of an analysis (even when main coder left the project)
- Working in same office helped
- Sprint retrospective very useful
  - Provided forum to even discuss conflicts between project members

Impact on Physics Result

How did the Result profit?

- *"Outcome could not have been better without using SCRUM"*
- Common tools, reusable tools enabled finished analysis
- Improved reproducibility (analysis was easily re-run on updated data set)

SCRUM in a Learning Environment

does SCRUM make sense working with students?

Experience gap

- Task breakdown difficult, if you are not experienced
- Daily meetings were essential to quickly remove obstacles
- SCRUM challenges Supervisors' assumptions on what instructions are needed and how to present them

Learning needs freedom

- Learning rarely straightforward $\leftrightarrow$ project tries to achieve a concrete goal
- Learning is difficult to plan
- Pressure to get something done at end of sprint
- This conflict is made visible by using SCRUM
- Allow time besides the Analysis Project to "just think"

Would you use SCRUM again?

What would you change?

- Yes, would use again (4/4)
- More homogenous team, more time (> 6months), longer sprints (2weeks?)
- Maybe a little bit more relaxed rules
- Better define / get rid of project owner
- Stricter timeboxing
- provide “free time” for people to work on side-projects or just “think”

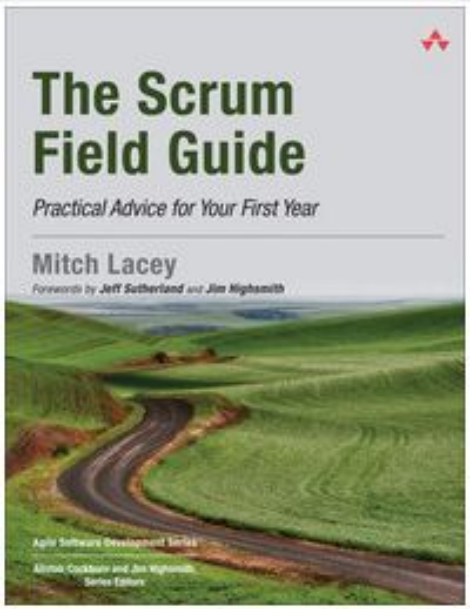
Conclusion: Successful Pilot Project

- Both analyses reached a state with interesting physics results
- SCRUM was appreciated and seen as useful
- Highlights the balance between teaching/learning and obtaining analysis results
- SCRUM is hard to master and needs practice
  - Need to understand the Values behind it
  - Beware of your gut-feeling, instead try to gather data and stay self-critical

How to get Started?

If you want to try it yourself

- Get some training! 2-day course offered by CERN (preferably by several team members)
- Read about it: There are a lot of good books (in the CERN library as well)
- Take your time, SCRUM is not easy to learn (> 3 months)
- Start with a fresh project



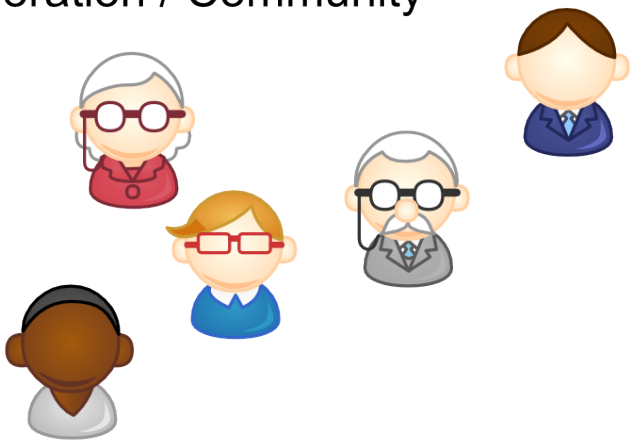
Tools

- git
- http://trello.com (integrates with git)
- JIRA/confluence

From Team to Community

Changing the questions you ask

Collaboration / Community



the Analysis

How does the measurement fit in the big picture?
What makes the analysis useful to the community?
How does our code help others to work with the results?

Team

Reproducible Experiment Platform (**Poster session A, booth # 25**)

Thank you!

BACKUP

Why did you choose SCRUM?

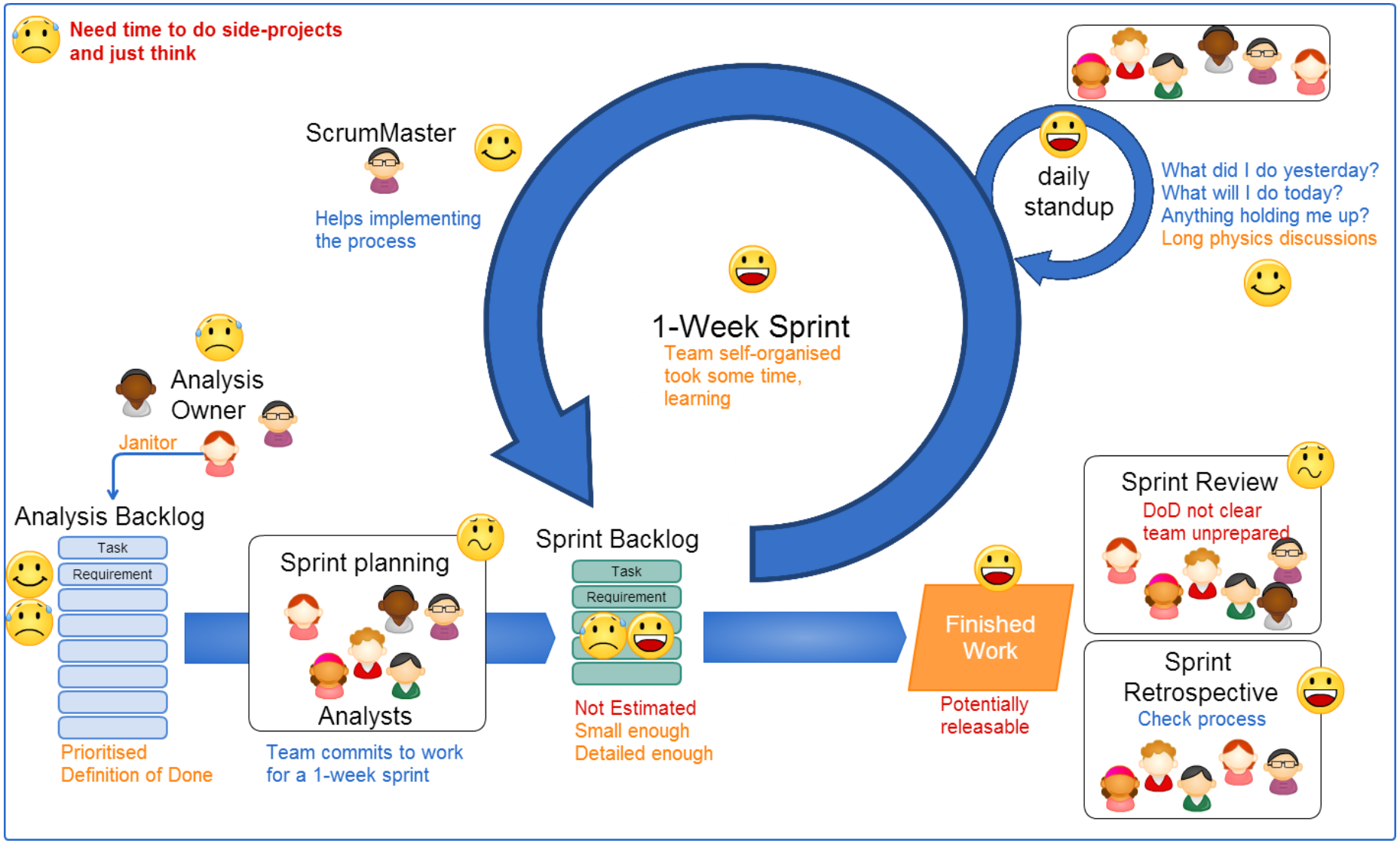
- Leading framework in software development – lot's of experience
- It fits our culture
- Lightweight and flexible
- Empirical process control
  - Stop theorising about processes.

    Start trying things and measure how they work
- SCRUM is a problem–detector

  (if you do it right)

How happy were we with SCRUM?

Summary extracted from survey data



What was most useful? Most problematic?

Most Useful:

- Backlog in Trello (4/4 mentions)
- Daily meeting (2/4 mentions)
- Regularity of sprint events (1 mention)

Most problematic:

- Following the rules too strictly (which we did not do)
- Team has little freedom to digress

What a Backlog does for you

Priorisation

- Focus on important points
- Tackle core problems first
- Confidently decide when it is enough

Task Breakdown

- Sparks discussions on what is needed
- Shows all the work (also the technical stuff)
- Keeps a logbook of what has been achieved and of appraoches that did not succeed

What Iterations do

Regular, frequent Feedback

- Inside the team (dailies, retrospective)
- From WG and Reviewers
- Establishes a rythm, removes deadline pressure

Allows to change plans

- Changes in analysis strategy have to happen as early as possible
- Iterations should be short

What the Team Commitment does

Shared ownership

- Problems are spotted early on when they are easy to correct
- Shared expertise
- Better maintainable code

Limitations

Keep boundary conditions in mind!

- Time: 12 weeks is very short $\Rightarrow$ pressure to quickly make progress on physics analysis
- 1 student leaving already after 8 weeks
- 3 Supervisors + 3 Students (balance?)
- Summie projects were not part of supervisors' main projects
- We were all beginners (one of us took a 2-day course in Scrum)

Technical Tools

What did the team use and build for themselves?

Used development tools

- CMake build environment
- GITHUB repository
- Interactive root sessions for more explorative tasks
- A virtual Pinboard to keep track of work

Analysis implemented as a set of compiled root executables

- shared between both analyses:
- BDT Training
- Data selection tool “secateur”
- Plotter
- Fitter
- Simple config files
- a few analysis specific scripts

Advice for Running a SCRUM Project

Try it out yourself!

- Stick to short sprints (1 week)
- Define clear tasks. Feature centered analysis:
  build the algorithm that will produce the plots / numbers
- Use extensive definition of done
- Learning technical skills: discuss code
- Use definition of ready for handover from supervisor to student
- Go out of your comfort zone to find right level of task decomposition (more details)
- Use task estimation/monitoring to learn about the team and the project
- Consciously leave space for side projects and “just thinking”

Evaluation: SCRUM Roles

- ProjectOwner
  - Unclear distinction between: ProjectOwner vs Supervisor
  - All three Supervisors acted as ProjectOwners
  - Generally role of ProjectOwner not clearly defined
    - How should he handle the backlog?
- ScrumMaster
  - ScrumMaster also acting as ProjectOwner
  - ScrumMaster interventions appreciated
- Team
  - Team needed time to self-organize, improvement over time visible

Evaluation: Artifacts

- Project Backlog
  - Trello tool was awesome to keep a visible backlog
  - Some feedback was divided:
  - priortisation not clear $\leftrightarrow$ prio not respected by team
  - “Prioritisation given to team to give them freedom”
  - Eternal items $\leftrightarrow$ dynamic, changing, never complete
- Sprint Backlog
  - Goal of sprint / project not well defined
  - Supervisors interferred a lot here, over time team took over
  - Items not estimated
  - Often items left after Sprint
- Potentially Releasable Increments
  - Not well defined
  - Completed items not structured/tested
  - Definition of Done not respected, was felt this would distract too much from work on physics

Evaluation: Events

- 1 week sprint worked very well
- Sprint Planning
  - It was clear WHAT to do
  - Not clear HOW to do it
  - Not enough task decomposition
  - Learning tasks difficult to plan
  - No Estimation done
- Daily meeting
  - worked very well
  - Sometimes deteriorated into too much physics discussion
  - But: physics discussion were appreciated
- Sprint Review
  - Definition of done not checked
  - Team often not well prepared
- Sprint Retrospective
  - Worked very well, had one situation where a student expressed his dissatisfaction with the discussion athmosphere, forum to discuss these issues was appreciated

Artifacts

- **Project backlog** : Prioritized and Estimated ToDo-List
- **Sprint backlog** : Items for current sprint, split into tasks
- Potentially releasable **analysis increment**
- **Definition of done**: when is a task finished?

Artifacts and Roles

Nuts and bolts of Scrum

Artifacts

- **Project backlog** : Prioritized and Estimated ToDo-List
- **Sprint backlog** : Items for current sprint, split into tasks
- Potentially releasable **analysis increment**
- **Definition of done**: when is a task finished?

Roles

- Self-organized analysis **team**
  - Everyone directly working on the analysis
  - Postdocs & Students
- **Analysis owner** (Moritz Karbach)
  - Keeps the Backlog up to date (vetoe on backlog changes)
  - Point of contact for working group and reviewers
- **Scrum master** (S.N.)
  - Helps the team to do Scrum
  - Vetoe on changes to process

What's going to happen?
Activities in Scrum

- Kick-off workshop
  - 2 (half)days workshop
  - Introducing physics topic
  - Setting up computing infrastructure
  - Scrum workshop

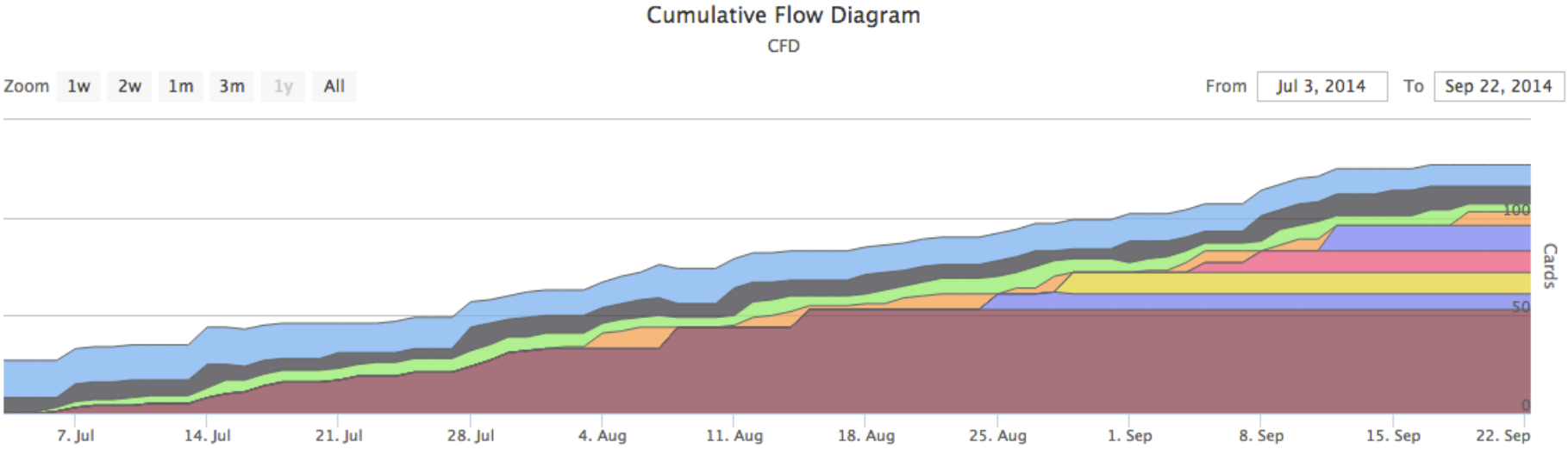
What's going to happen?
Activities in Scrum

- Kick-off workshop
  - 2 (half)days workshop
  - Introducing physics topic
  - Setting up computing infrastructure
  - Scrum workshop

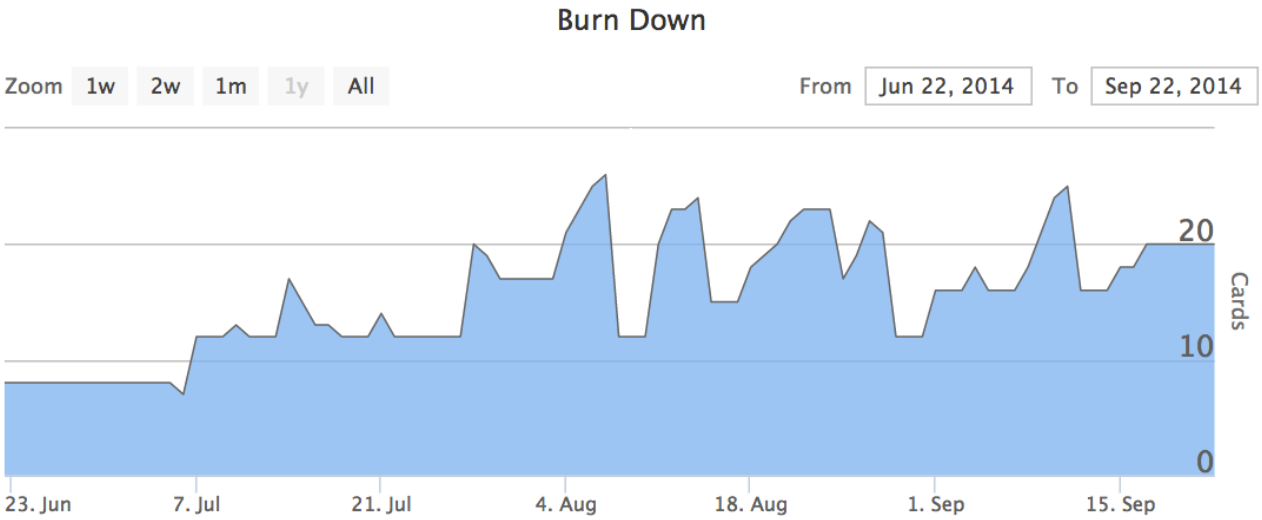
- 1-week Sprints:
  - Team working together in a common office
  - **Sprint planning workshop** Timeboxed: $\sim 2h\,(< 8h)$
    Discuss next steps, create sprint backlog
  - **Daily coffee** $< 15min$
    Super-Quick status report
  - **Sprint review** $\sim 1h\,(< 4)$
    Demo working analysis software (no slides)
  - **Sprint retrospective** $\sim 1h\,(< 3)$:
    Reflecting on HOW the team worked, discussing improvements

Trello Board Usage

Analysis done after project with ↪ ollertapp.com



- 121 cards
- 82% completed



Why is it Important to Track Progress?

- Being stuck indicates problems that need attention
- We often tend to forget what we already have accomplished
- Acknowledging progress is essential for motivation

- Relentlessly collective code ownership (everybody in a WG is encouraged to contribute to every analysis)
- Enable future analysers to easily repeat your analysis
- Enable micro-contributions
- Shift perspective from the plots to the tools that produce the plots
- `make analysis` ($\hookrightarrow$ example from bioinformatics)

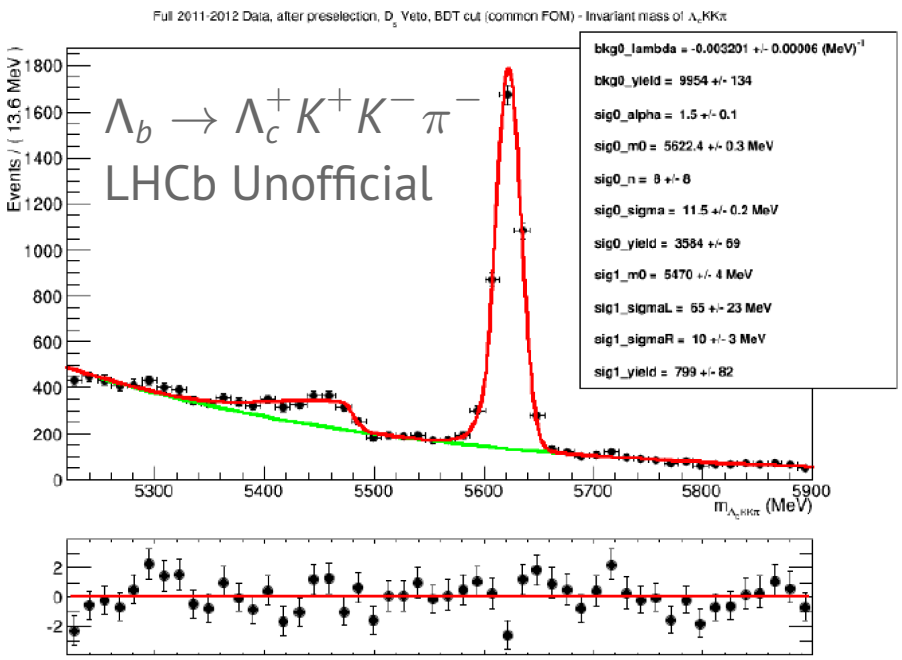
What Comes Out of a Sprint?
Potentially Releasable Results

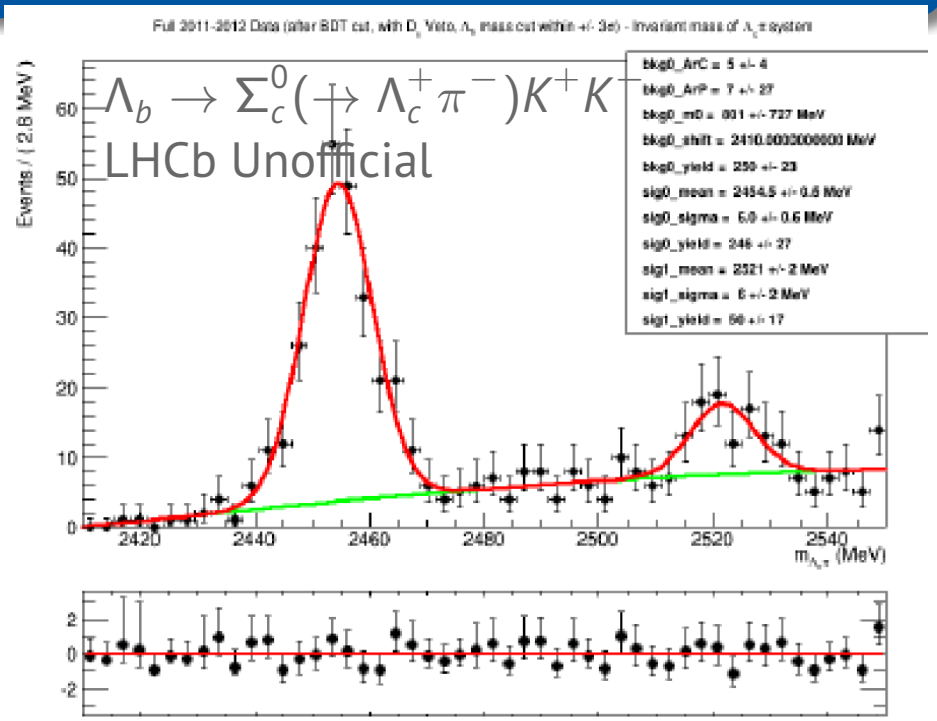
To achieve a potenitally releasable results:

- Clear Definition of Done (it's difficult but worth it)
- Include Reviewers from Day 1 $\rightarrow$ Sprint Reviews
  - Don't mix Sprint Reviews with WG meetings!!!
  - Use the prioritised backlog to decide what to do next and when to stop

Analysis Results: $\Lambda_b \to \Lambda_c^+ K^+ K^- \pi^-$

Search for new decay modes





Potentially 5 first observations

Decay (fully reconstructed)	Signal yield	Bkg. yield	FOM
$\Lambda_b^0 \to \Lambda_c^+ K^- K^+ \pi^-$,	3584 ± 69	685 ± 26	54.8
$\Lambda_b^0 \to \Sigma_c(2455)^0 K^- K^+$, $\Sigma_c(2455)^0 \to \Lambda_c^+ \pi^-$	246 ± 27	47 ± 7	14.4
$\Lambda_b^0 \to \Sigma_c(2520)^0 K^- K^+$, $\Sigma_c(2520)^0 \to \Lambda_c^+ \pi^-$	50 ± 17	96 ± 10	4.1
$\Lambda_b^0 \to \Lambda_c^+ K^* K^-$, $K^* \to K \pi$	2267 ± 67	220 ± 15	45.5
$\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$, $\bar{D}^0 \to K^+ \pi^-$	113 ± 12	25 ± 5	9.6
Decay (partially reconstructed)	Signal yield	Bkg. yield	FOM
$\Lambda_b^0 \to \Sigma_c(2455)^+ K^- K^+ \pi^-$, $\Sigma_c(2455)^+ \to \Lambda_c^+ \pi^0$	799 ± 82	4833 ± 70	10.6
$\Lambda_b^0 \to \Sigma_c^0 \bar{D}^0 K^-$, $\bar{D}^0 \to K^+ \pi^-$, $\Sigma_c^0 \to \Lambda_c^+ \pi^0$	274 ± 32	140 ± 12	13.5
$\Lambda_b^0 \to \Lambda_c^+ D^*(2007)^0 K^-$, $\bar{D}^*(2007)^0 \to \bar{D}^0 \pi^0$	130 ± 49	323 ± 18	6.1

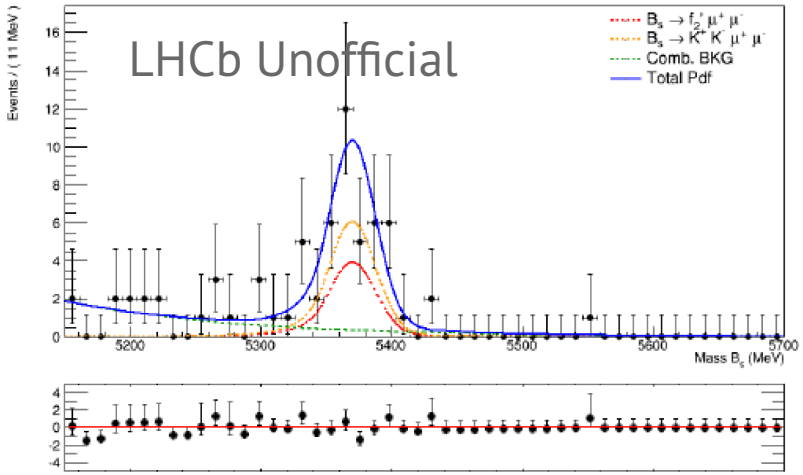
Summer student reports:

- $\hookrightarrow$ K. Heijhoff (BDT)
- $\hookrightarrow$ V. Andreev
  (Fit, 2-body subsystems)

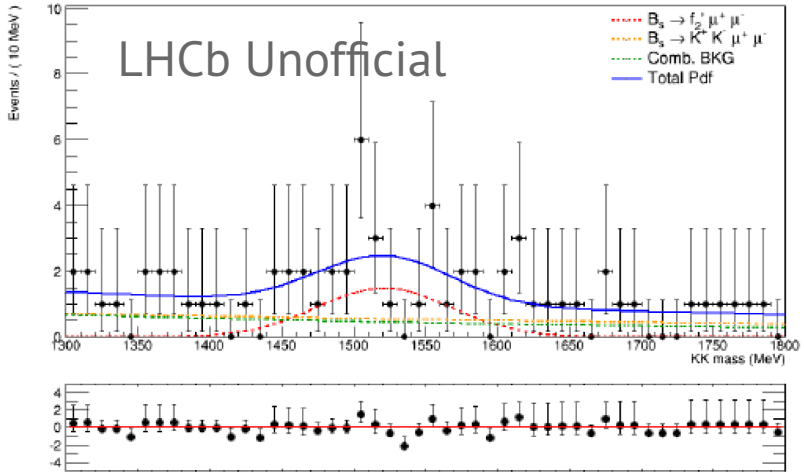
The Rare Decay $B_s \to f_2(1525)\mu\mu$

with control channel $B_s \to f_2(1525)J/\psi$

2-dim fit

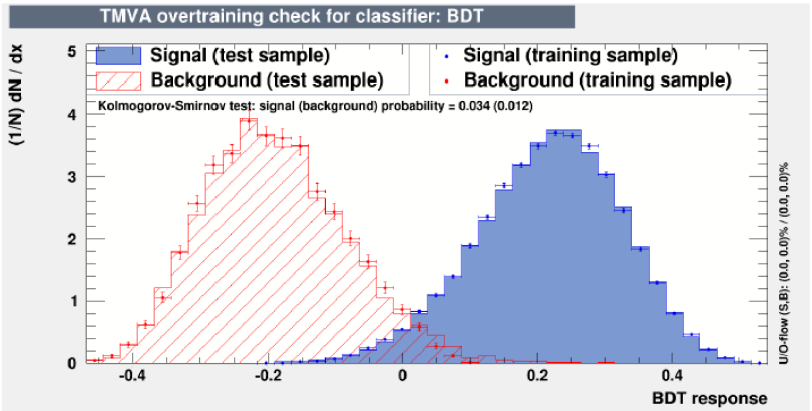


(a) Projection on B_s mass



(b) Projection on KK mass

BDT trained on signal MC and B_s sideb. bkg



- Preliminary efficiency corrected yield in agreement with SM prediction [PRD83 (2011)034034]
- Report: ↪ D. Berninghoff