

# Sherpa W+Jet Profiling

---

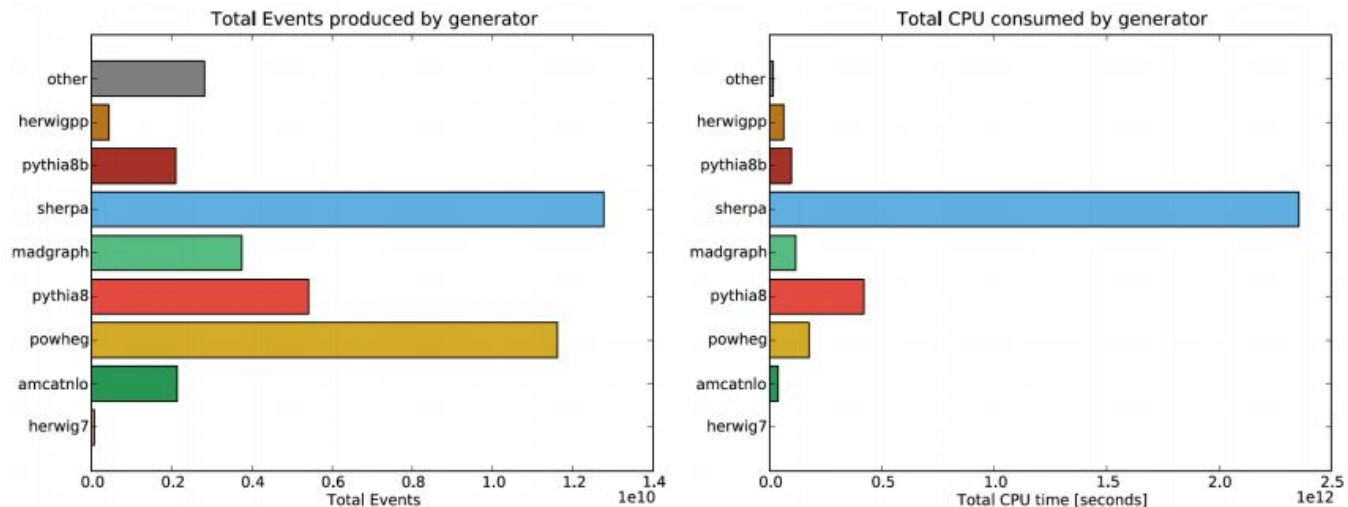
Tim Martin, Warwick  
June 19<sup>th</sup> 2020



# Intro

- Noted in kickoff meeting that the largest CPU draw in ATLAS Event Generation comes from Sherpa SM processes. (CMS dominated by MadGraph, to a lesser extent)
- Investigate where this comes from, look for potential improvements.

→ right plot: most CPU spent on high-precision calculations for  $V + 0, 1, 2j@NLO+3, 4j@LO$  and  $t\bar{t} + 0, 1j@NLO+2, 3, 4j@LO$



# Initial State

- Everything compiled out-the-box at O2
- Full ATLAS-representative  $W$ +Jets setup provided by Marek Schoenherr
  - **$W+0,1,2j@NLO+3,4,5j@LO$**
  - Including approximate virtual corrections and reweightings to different PDFs and scales
- Running 500 events EvGen
- Single-core
- Total time: 19,876 s (around 5h 30m)

# Software and PC Details

- Local compilations of
  - Sherpa 2.2.8
  - OpenLoops 2.1.1
  - LHAPDF 6.2.3
  - HepMC 3.2.0
  - Intel(R) VTune(TM) Profiler 2020 (build 605129)
- Local software stack provides
  - gcc 5.5.0
  - Intel(R) icc 16.0.3 20160415
  - cmake 3.12.2
  - sqlite 3.24.0
  - root 6.1

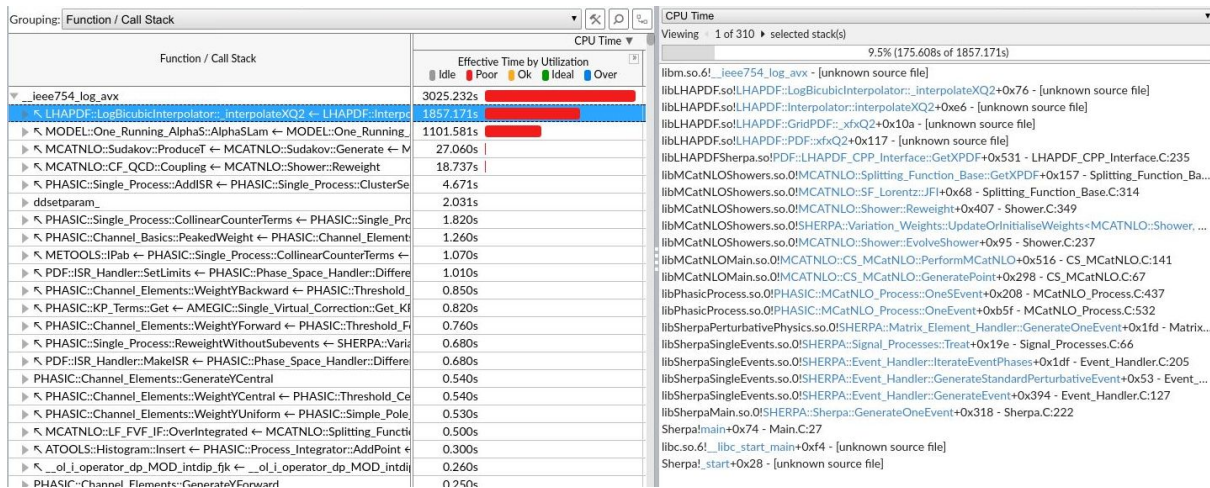
Profiling data collected by vtune running in userspace.  
Visualisations of vtune database via [flamegraph](#).

## CPU Details

```
vendor_id : GenuineIntel
cpu family      : 6
model          : 158
model name     : Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz
stepping      : 9
microcode     : 0xd6
cpu MHz       : 900.000
cache size    : 8192 KB
```

# Vanilla CPU #1: log

- Unsurprisingly large call on transcendental functions. Two main callees
  - > LHAPDF::LogBicubicInterpolator::\_interpolateXQ2 -> LHAPDF::Interpolator::\_interpolateXQ2 -> LHAPDF::GridPDF::\_xfxQ2 -> LHAPDF::PDF::\_xfxQ2 -> PDF::LHAPDF\_CPP\_Interface::GetXPDF 1857.171s 0ms 0usec libLHAPDF.so LHAPDF::LogBicubicInterpolator::\_interpolateXQ2 (LHAPDF::KnotArray1F const&, double, unsigned long, double, unsigned long) const [Unknown] 0x453b0
  - > MODEL::One\_Running\_AlphaS::AlphaSLam -> MODEL::One\_Running\_AlphaS::operator() 1101.581s 0ms 0usec libModelMain.so.0 MODEL::One\_Running\_AlphaS::AlphaSLam(double, int) Running\_AlphaS.C 0x221b0



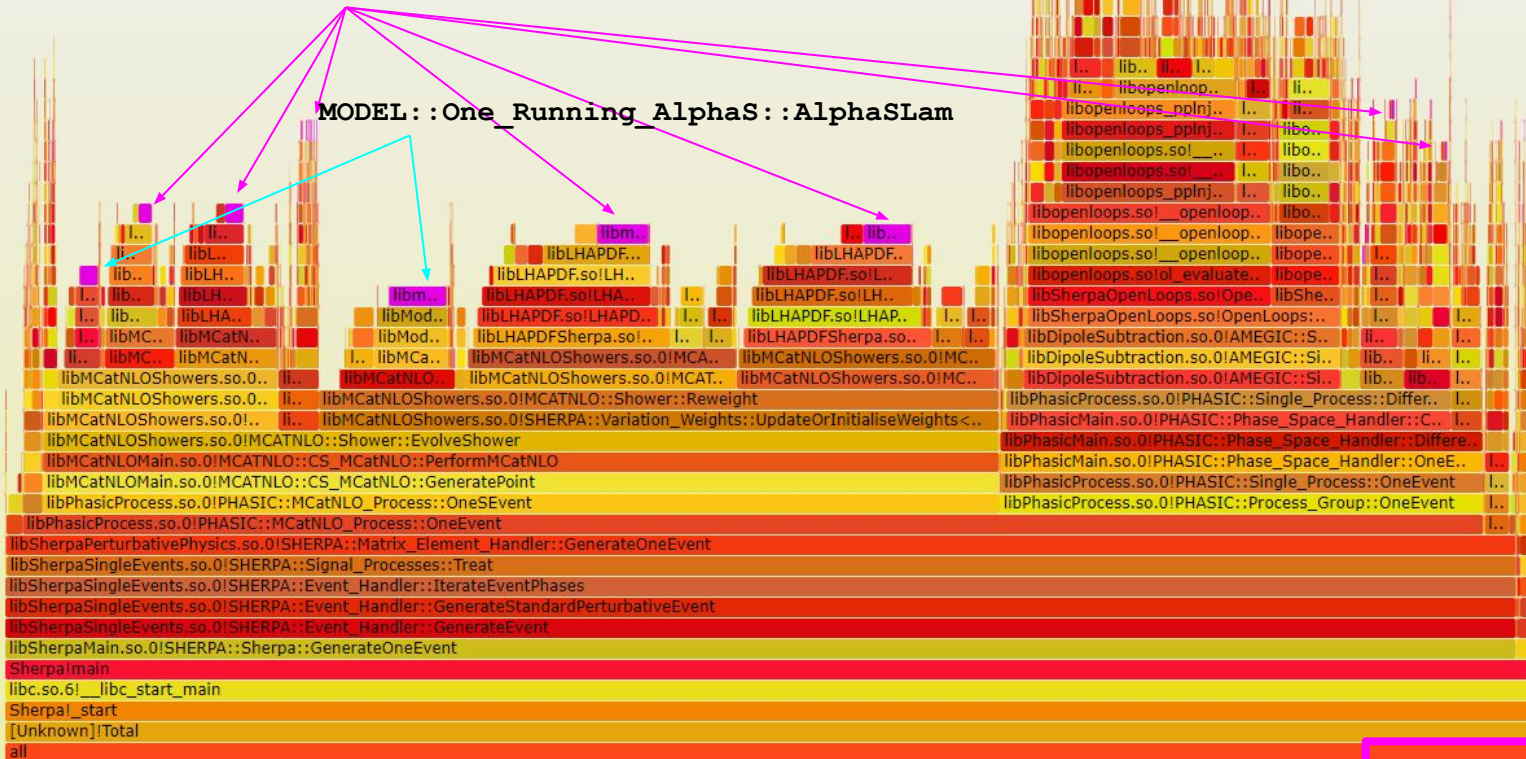
# Calls to `__ieee754_log_avx`

Flame Graph

Reset Search

`LHAPDF::LogBicubicInterpolator::_interpolateXQ2`

`MODEL::One_Running_AlphaS::AlphaSLam`

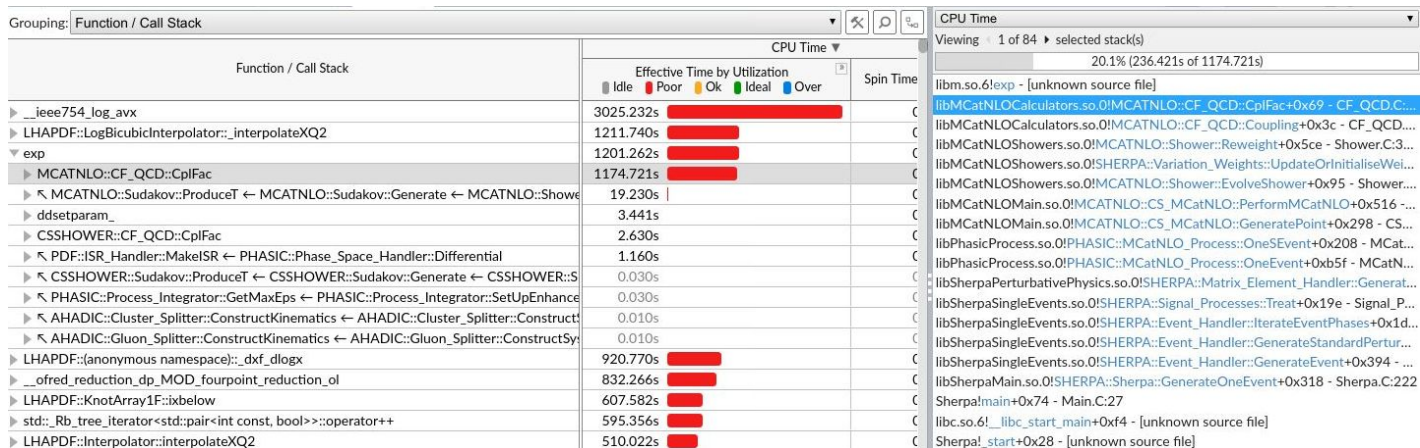


Matched: 15.1%



# Vanilla CPU #3: **exp**

- Another transcendental takes the third slot at 1,174 s
- `MCATNLO::CF_QCD::CplFac 1174.721s 0ms 0usec libMCatNLOCalculators.so.0`  
`MCATNLO::CF_QCD::CplFac(double const&) const CF_QCD.C 0xb420`



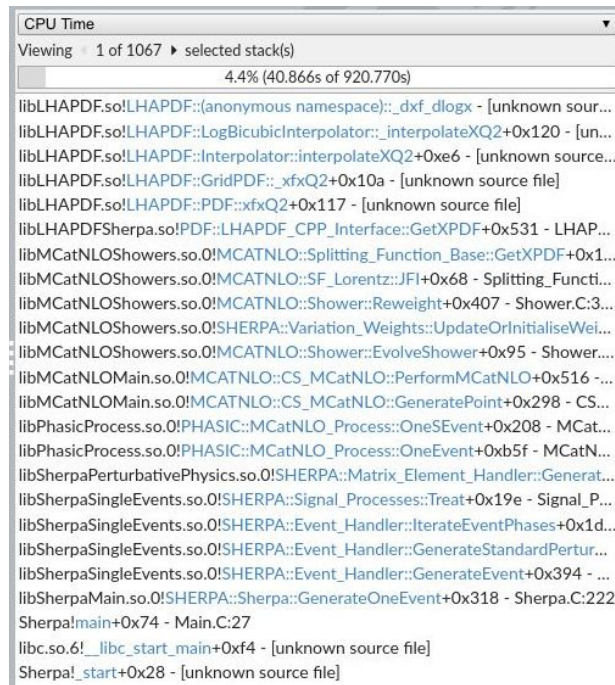
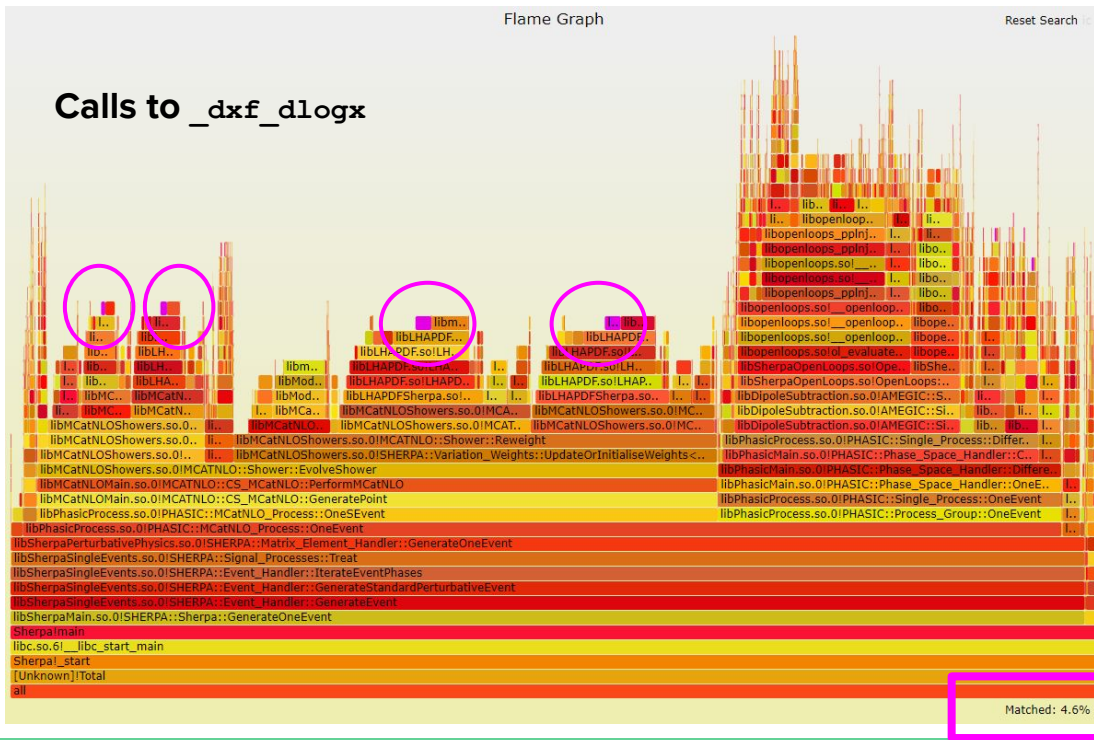
```
191 double CF_QCD::CplFac(const double &scale) const
192 {
193     if (m_kfmode==-1) return 1.0;
194     if (m_kfmode==0) return m_cplfac;
195     One_Running_AlphaS * const as = (p_altcpl) ? p_altcpl : p_cpl->GetAs();
196     double nf=as->Nf(scale);
197     > double kfac=exp(-(67.0-3.0*sqr(M_PI)-10.0/3.0*nf)/(33.0-2.0*nf));
198     return m_cplfac*kfac;
199 }
200
```





# Vanilla CPU #4: **\_dxf\_dlogx**

- Like #1, #4 is a child of #2 **\_interpolateXQ2**
- LHAPDF::LogBicubicInterpolator::\_interpolateXQ2 -> LHAPDF::Interpolator::\_interpolateXQ2 -> LHAPDF::GridPDF::\_xfxQ2 -> LHAPDF::PDF::\_xfxQ2 -> PDF::LHAPDF\_CPP\_Interface::GetXPDF 920.770s



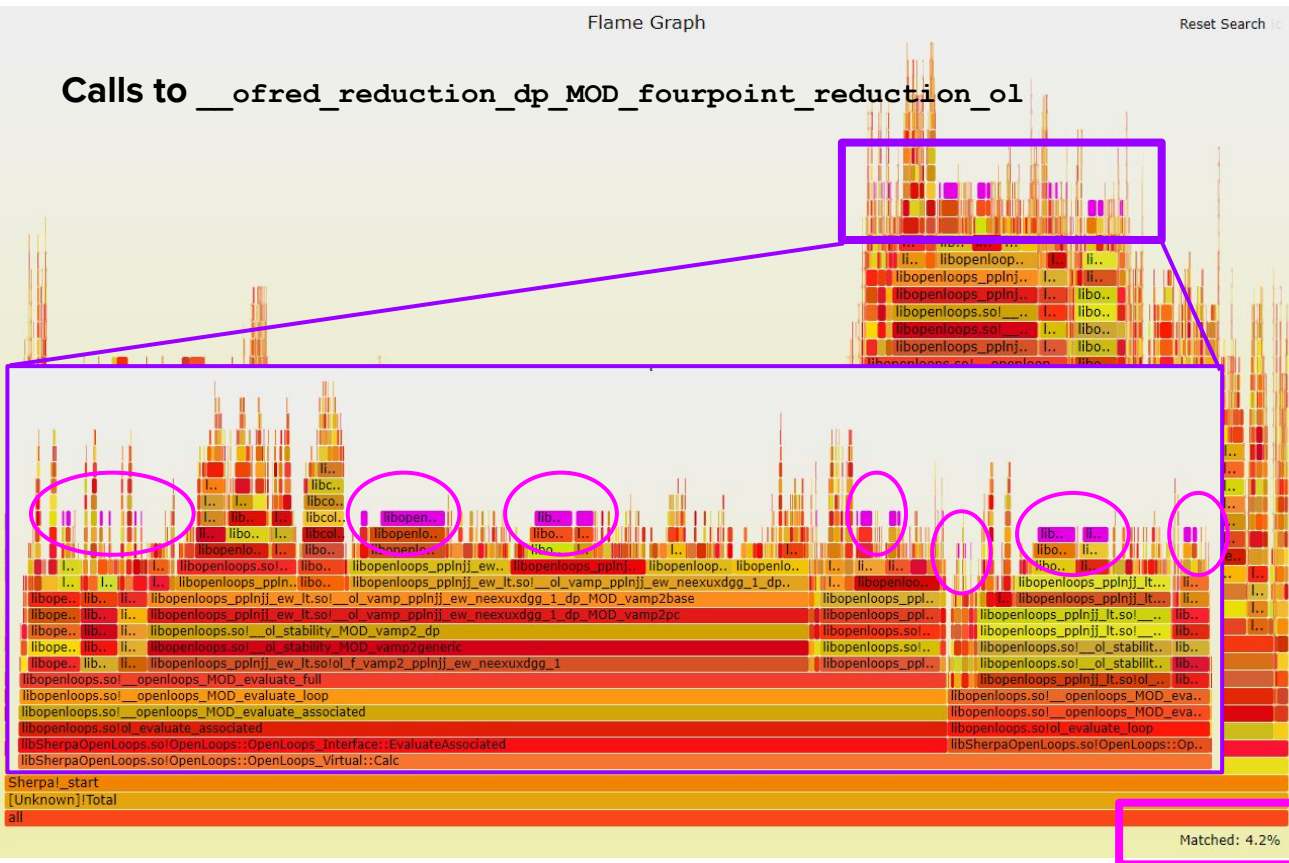
# Vanilla CPU #5: **ofred\_reduction\_dp\_MOD\_fourpoint\_reduction\_ol**

- Top single CPU calling coming from Openloops at 832 s

Flame Graph

Reset Search

## Calls to **\_\_ofred\_reduction\_dp\_MOD\_fourpoint\_reduction\_ol**

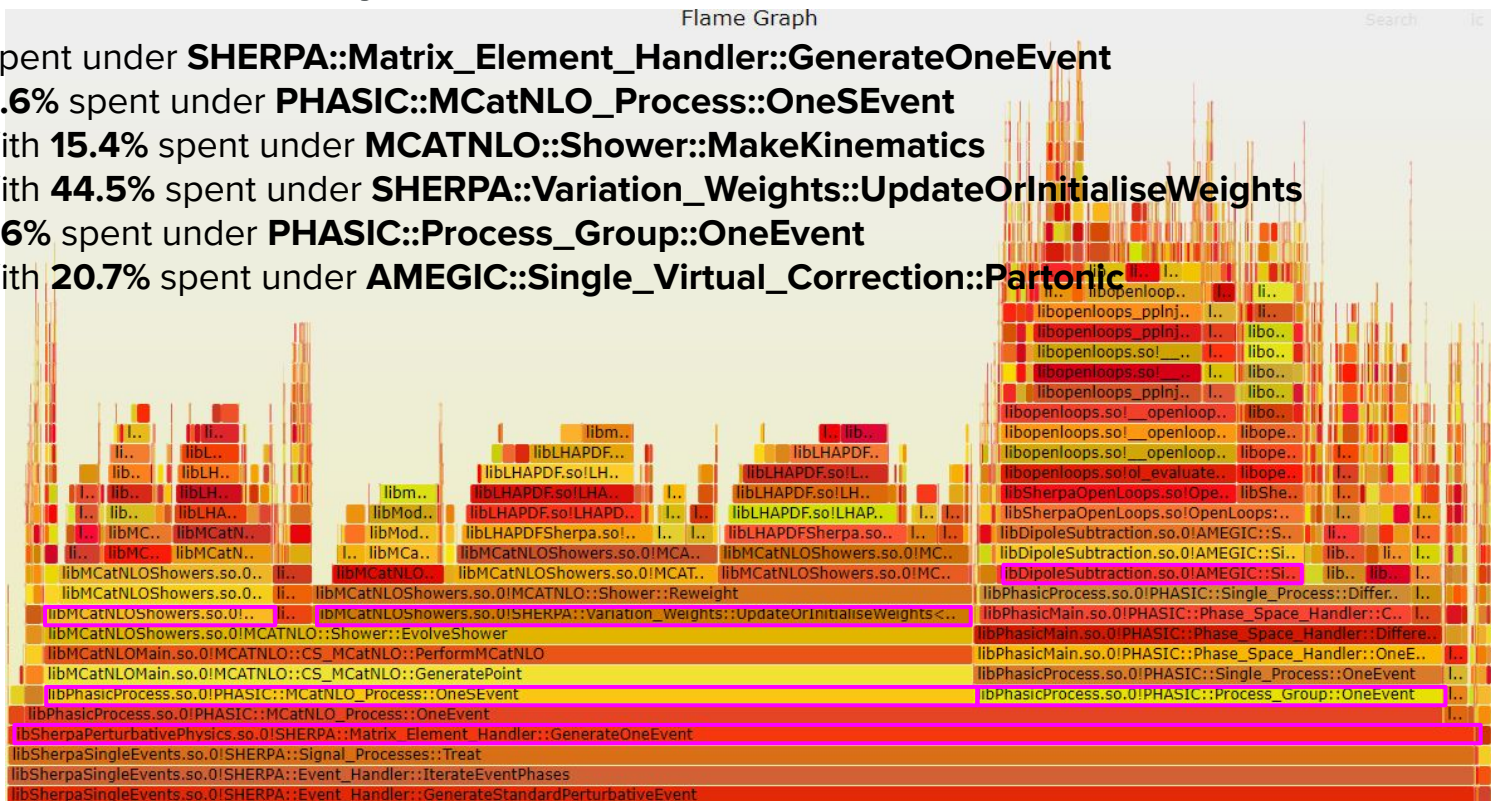


```
CPU Time
Viewing 1 of 936 selected stack(s)
1.6% (6.683s of 428.928s)
libopenloops.so!__ofred_reduction_dp_MOD_fourpoint_reduction_ol - [unknown sou...
libopenloops.so!__ofred_reduction_dp_MOD_off_4pt_red+0x3f3 - [unknown source ...
libopenloops.so!__ofred_reduction_dp_MOD_hotf_4pt_red+0x248 - [unknown sourc...
libopenloops.pplnjj.It.so!_ol_vamp_1.pplnjj_neexuxdgg_1_dp_MOD_vamp_1+0x36...
libopenloops.pplnjj.It.so!_ol_vamp_pplnjj_neexuxdgg_1_dp_MOD_vamp2+0x109...
libopenloops.pplnjj.It.so!_ol_vamp_pplnjj_neexuxdgg_1_dp_MOD_vamp2pc+0x109...
libopenloops.so!_ol_stability_MOD_vamp2_dp+0x212 - [unknown source file]
libopenloops.so!_ol_stability_MOD_vamp2generic+0x14bc - [unknown source file]
libopenloops.pplnjj.It.so!ol_f_vamp2_pplnjj_neexuxdgg_1+0x142 - [unknown source ...
libopenloops.so!__openloops_MOD_evaluate_full+0x119a - [unknown source file]
libopenloops.so!__openloops_MOD_evaluate_loop+0xb - [unknown source file]
libopenloops.so!ol_evaluate_loop+0x261 - [unknown source file]
libSherpaOpenLoops.so!OpenLoops::OpenLoops_Interface::EvaluateLoop+0xfcd - Ope...
libSherpaOpenLoops.so!OpenLoops::OpenLoops_Virtual::Calc+0x176 - OpenLoops_V...
libDipoleSubtraction.so!AMEGIC::Single_Virtual_Correction::operator()+0x1033 - Si...
libDipoleSubtraction.so!AMEGIC::Single_Virtual_Correction::DSigma+0x3b4 - Singl...
libDipoleSubtraction.so!AMEGIC::Single_Virtual_Correction::Partonic+0x63 - Singl...
libPhasicProcess.so!PHASIC::Single_Process::Differential+0x1207 - Single_Proces...
libPhasicMain.so!PHASIC::Phase_Space_Handler::CalculateME+0x18 - Phase_Space...
libPhasicMain.so!PHASIC::Phase_Space_Handler::Differential+0x55a - Phase_Space...
libPhasicMain.so!PHASIC::Phase_Space_Handler::OneEvent+0x44 - Phase_Space_H...
libPhasicProcess.so!PHASIC::Single_Process::OneEvent+0x4d - Single_Process:C:58
libPhasicProcess.so!PHASIC::Process_Group::OneEvent+0x152 - Process_Group:C:48
libPhasicProcess.so!PHASIC::MCAtnLO_Process::OneEvent+0x290 - MCAtnLO_Pro...
libSherpaPerturbativePhysics.so!OSHERPA::Matrix_Element_Handler::GenerateOneE...
libSherpaSingleEvents.so!OSHERPA::Signal_Processes::Treat+0x19e - Signal_Proces...
libSherpaSingleEvents.so!OSHERPA::Event_Handler::IterateEventPhases+0x1df - Eve...
libSherpaSingleEvents.so!OSHERPA::Event_Handler::GenerateStandardPerturbativeE...
libSherpaSingleEvents.so!OSHERPA::Event_Handler::GenerateEvent+0x394 - Event_...
libSherpaMain.so!OSHERPA::Sherpa::GenerateOneEvent+0x318 - Sherpa.C:222
Sherpa/main+0x74 - Main.C:27
```

Matched: 4.2%

# Vanilla O2 Summary

1. **98%** of time spent under **SHERPA::Matrix\_Element\_Handler::GenerateOneEvent**
  - a. With **62.6%** spent under **PHASIC::MCatNLO\_Process::OneSEvent**
    - i. With **15.4%** spent under **MCATNLO::Shower::MakeKinematics**
    - ii. With **44.5%** spent under **SHERPA::Variation\_Weights::UpdateOrInitialiseWeights**
  - b. With **31.6%** spent under **PHASIC::Process\_Group::OneEvent**
    - i. With **20.7%** spent under **AMEGIC::Single\_Virtual\_Correction::Partonic**



- **AMEGIC, OpenLoops: Deep call-stacks, resource usage spread among many calls.**
- **LHAPDF: Shallow call-stack, large CPU cost from transcendental functions.**
- **MCatNLO: Somewhere in the middle between these two extremes.**

# Compile Time Optimisations

- Multiple easy “**slot in**” optimisation strategies tried.
  - **Memory allocation**: preload Google’s TCMALLOC
  - **Optimisation** level: O2 vs. O3
  - Beyond O3: **Minimum architecture** (msse4.2) and **unsafe maths** optimisations
  - **Link Time Optimisation**
  - Fully **static** single-process builds (**not managed successfully...**)
  - Use of Intel icc compiler.
    - O3, minimum architecture and unsafe maths flags with icc
- Optimisations were tried only on Sherpa and on Sherpa+LHAPDF. In some cases this resulted in a strange regression.
- TCMALLOC and VTune did not play nice.
- Working set of icc flags usable with both LHAPDF and Sherpa took some iterations.
- Have not checked so far if physics were impacted by unsafe maths.
- ***See backup slides for full details.***

# Optimisation Timings

Optimisation	On Sherpa	On Sherpa+LHAPDF
None (O2)	19,876 s	N/A
Memory Allocation	19,505 s (-2%)	~83,463 s (!)
O3	19,506 s (-2%)	20,280 s (+2%)
Architecture & Unsafe Maths	20,183 s (+2%)	~82,785 s (!)
Link Time Optimisation	20,371 s (+2%)	~80,935 s (!)
Intel compiler		18,122 s (-9%)
Intel compiler + O3, Arch, Maths		18,598 s (-6%)

Intel savings look to come primarily through faster maths. C.f.

```
(gcc) __ieee754_log_avx = 3,025 s
```

```
(icc) __libm_log_19 = 956 s
```

See also Manuel's slides from

[yesterday's ECHP Reco meeting](#).

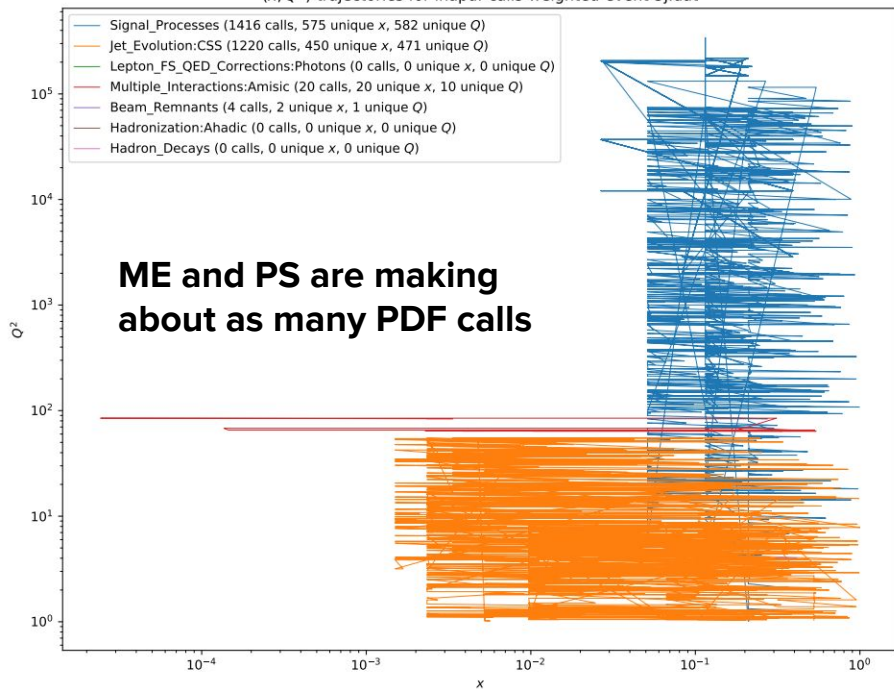
Vectorisation can save much more.



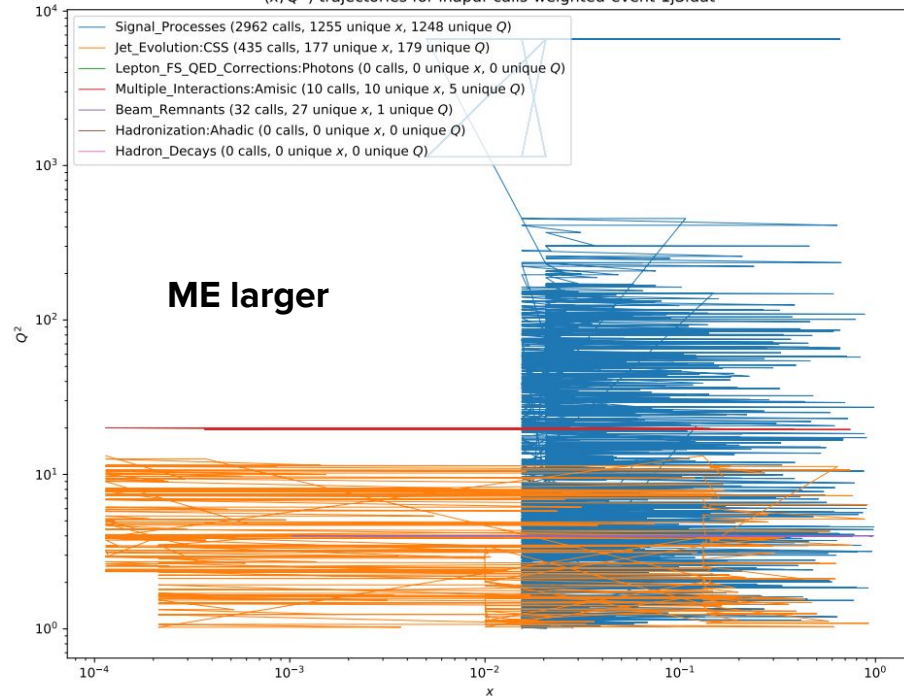
# Investigation into PDF Evaluation Calls

- Single event dumps of calls by Sherpa produced by Marek
  - $\langle \text{flavour} \rangle, \langle x \rangle, \langle Q^2 \rangle$  for some different types of events (pp $\rightarrow$ W+0j S/H, 1j S/H, 5j LO)
- Visualised by Andy

$(x, Q^2)$  trajectories for lhpdf-calls-weighted-event-5j.dat



$(x, Q^2)$  trajectories for lhpdf-calls-weighted-event-1jS.dat





# Caching in LHAPDF

**Tested:**  
**With 6.2.4beta1**  
**With 6.3.0beta1**  
**With 6.3.0beta3**

- Investigating caching of values in LHAPDF
  - Some work carried out previously by Dima Konstantinov and Grigorii Latyshev
  - Additional work by Andy Buckley
- Iterations with implementing small thread-local caches (**current size: 4**).
  - **Avoid re-interpolation** when the same value is requested multiple times in a single event.
- May also benefit from revisiting how Sherpa structures its LHAPDF calls.
- May also benefit from caching over different grids in LHAPDF.

Version	% in LHAPDF	Total Time
6.2.4	47.3%	19876 s
6.2.4beta1	43.9%	20235 s
6.3.0beta1	44.3%	20505 s
6.3.0beta2	43.7%	19634 s



# Summary

- Sherpa W+jets is a known huge CPU consumer on ATLAS
- CPU savings can be made by use of Intel's maths library, more may be possible via vectorized maths libraries.
- Better integration with LHAPDF could yield larger saving still.
- Flamegraphs at <https://cernbox.cern.ch/index.php/s/9OK2W17XELp6Gmt>

Backup

---

# Optimisations #1: malloc

- `malloc` consumes 1.7% and `operator new()` consumes 0.9%
- [TCMalloc : Thread-Caching Malloc](#)
- *“TCMalloc is Google's customized implementation of C's malloc() and C++'s operator new used for memory allocation within our C and C++ code. TCMalloc is a fast, multi-threaded malloc implementation.”*
- Around 2.3x faster than `malloc`, used by ATLAS
- Unfortunately, TCMalloc did not play nice with VTune. Just timed with **time**
- **Sherpa + TCMalloc: 19,505 s (-2%)**



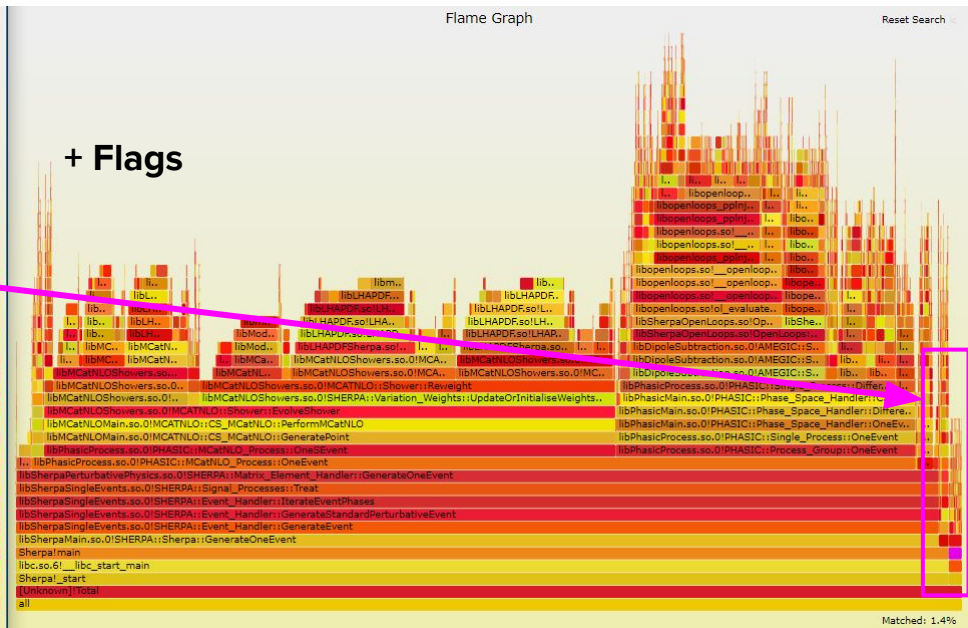
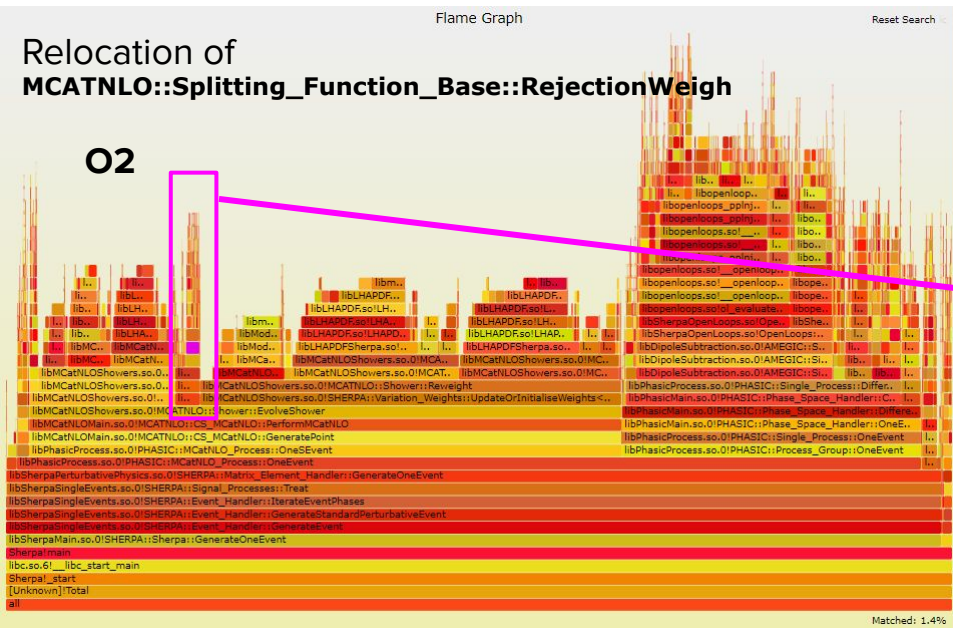
# Optimisations #3: Architecture & Unsafe Maths

- **-msse4.2** specifies to use the SSE4.2 CPU instruction set extension.
  - Available since Nehalem (initial core i5, i7), November 2008
  - LHCb reported having fully switched over to this instruction set.
  - ATLAS has recent experience of some grid nodes and user machines *still* not supporting it.
- **-ffast-math** applies a bunch of “unsafe”\* operations, not applied by -O.
  - **-fno-trapping-math, -fno-signaling-nans**: User **cannot trap /0** or overflow.
  - **-funsafe-math-optimizations**: *This mode enables optimizations that allow arbitrary reassociations and transformations with no accuracy guarantees. Due to roundoff errors the associative law of algebra do not necessary hold for floating point numbers and thus **expressions like  $(x + y) + z$  are not necessary equal to  $x + (y + z)$***
  - **-ffinite-math-only**: *Assume that there will **never be NaNs or +/-Infs***
  - **-fno-errno-math**: *Disables setting of the errno variable as required by C89/C99 on calling math library routines.*
  - **-fno-rounding-math**: *IEEE has four rounding modes. This flag assumes that the rounding mode is round to nearest.*
  - **-fcx-limited-range**: *Causes the **range reduction step to be omitted when performing complex division**. This uses  $a / b = ((ar*br + ai*bi)/t) + i((ai*br - ar*bi)/t)$  with  $t = br*br + bi*bi$  and might not work well on arbitrary ranges of the inputs.*
  - **-fno-signed-zeros**: **Removes the ability to have signed 0**

# Optimisations #3: Architecture & Unsafe Maths

- Again, few changes to the call graph.
- Physics output not yet checked!
- **Sherpa -mste4.2 -ffast-math: 20,839 s (+5%)**

Needs re-running, other cores were in use...

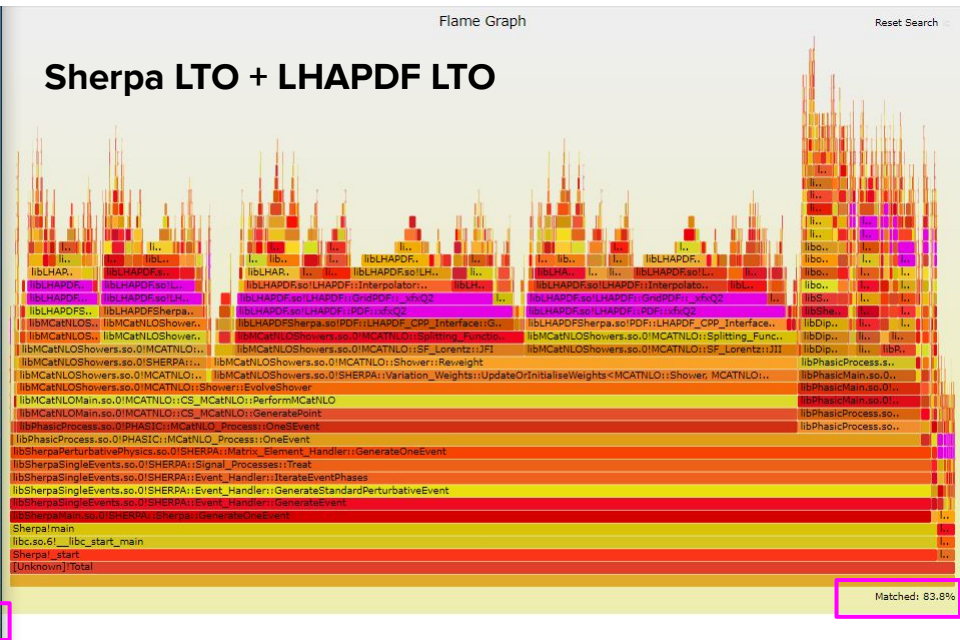
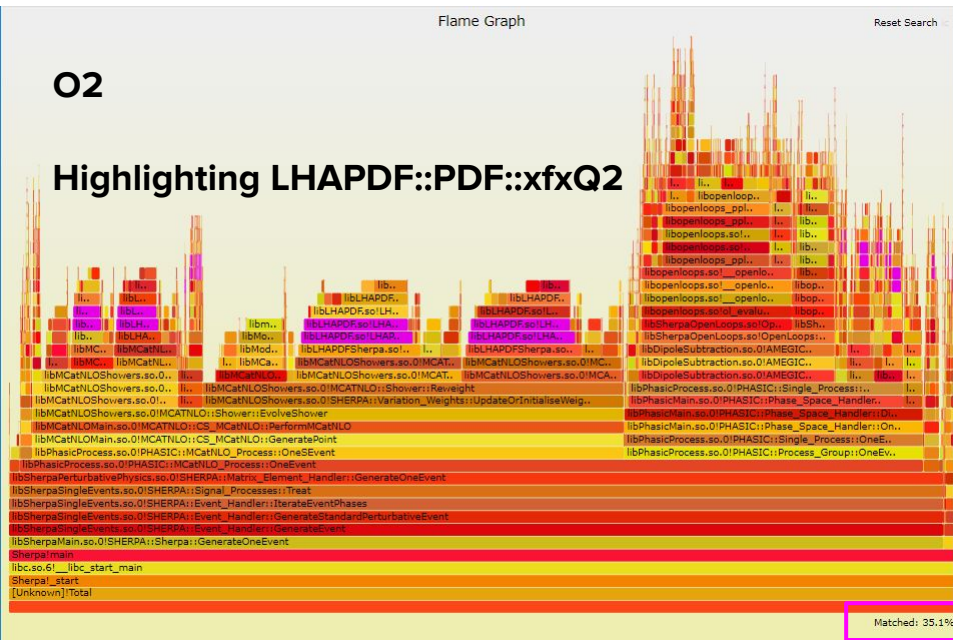






# Other permutations:

- Trying to spread the flags also to LHAPDF encountered a serious regression!
  - O2: **40 s/event**
  - Sherpa O3 + LHAPDF O3 + OPENLOOPS fortran O3: **40 s/event**
  - Sherpa + TCMalloc , LHAPDF + TCMalloc: **167 s/event (!!!)**
  - Sherpa LTO + LHAPDF LTO: **161 s/event (!!!), 150 s/event (!!!)**
  - Sherpa Fast Maths flags + LHAPDF Fast Maths flags: **152 s/event (!!!)**



# Other permutations:

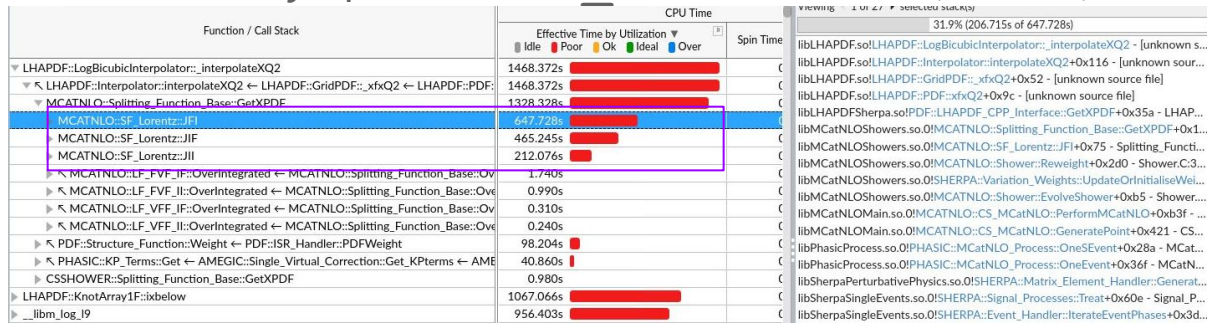
- Trying a fully static build of Sherpa.
  - Tried to link Sherpa against minimal set of dependencies.
  - Build / combine static `.a` files for all required libraries.
  - [Link them all together statically](#)
- Managed to make a giant binary.
- Couldn't get it to work... segfaults...

# Intel

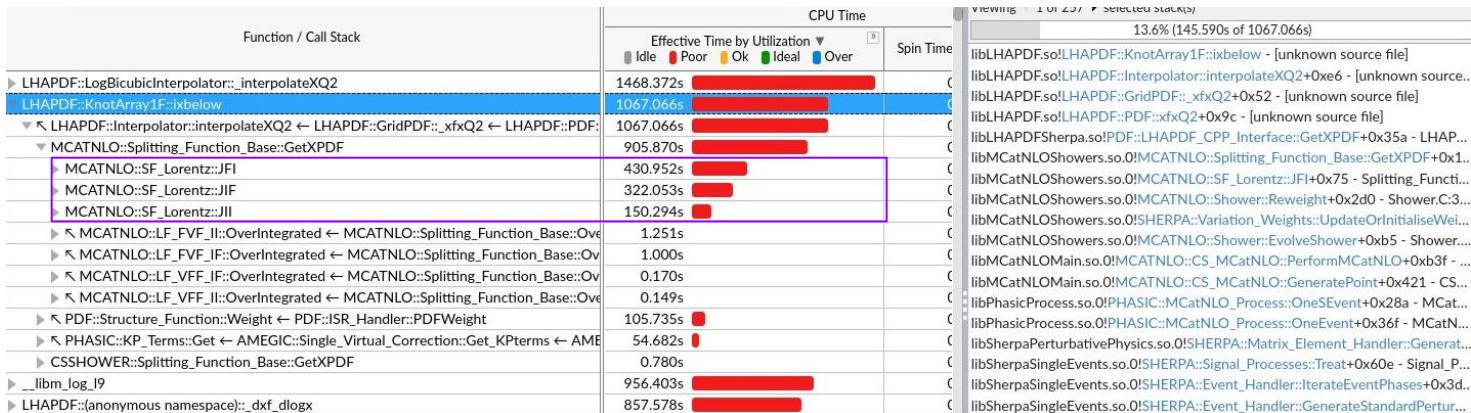
- Proprietary compiler, but licencing may be available through CERN
- **Note:** ATLAS preload the intel maths libraries
- Using `icc 16.0.3 20160415` licenced by Warwick
- Total time: **18,025s**
  - **10% faster than GCC**

# CPU #1: LHAPDF::LogBicubicInterpolator::\_interpolateXQ2

- Predominantly split over **SF\_Lorentz::JFI, JIF, JII**

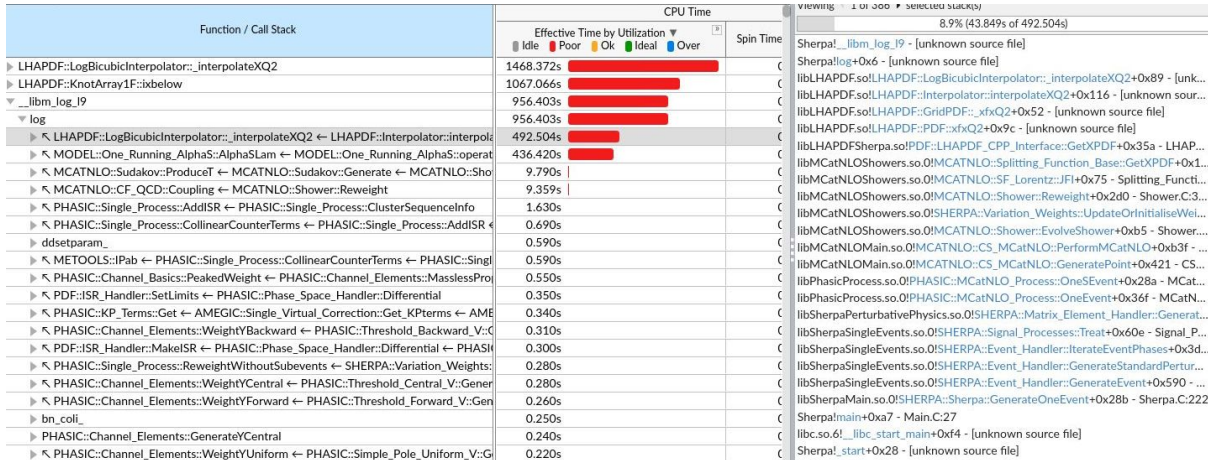


# CPU #2: LHAPDF::KnotArray1F::ixbelow



# CPU #3: `__libm_log_I9`

- Log takes the #3 slot, calls from
  - LHAPDF::Interpolator::interpolateXQ2
  - MODEL::One\_Running\_AlphaS::operator()



# CPU #4: \_dxf\_dlogx

Function / Call Stack	CPU Time				Spin Time
	Effective Time by Utilization				
	Idle	Poor	Ok	Ideal	
▶ LHAPDF::LogBicubicInterpolator::_interpolateXQ2	1468.372s				(
▶ LHAPDF::KnotArray1F::ixbelow	1067.066s				(
▶ __libm_log_l9	956.403s				(
▼ LHAPDF:(anonymous namespace)::_dxf_dlogx	857.578s				(
▼ LHAPDF::LogBicubicInterpolator::_interpolateXQ2 ← LHAPDF::Interpolator::interpolate	857.578s				(
▶ MCATNLO::Splitting_Function_Base::GetXPDF	739.485s				(
▶ PDF::Structure_Function::Weight ← PDF::ISR_Handler::PDFWeight	71.891s				(
▶ PHASIC::KP_Terms::Get ← AMEGIC::Single_Virtual_Correction::Get_KPterms ← AMEGIC::Single_Virtual_Correction::Get_KPterms	45.662s				(
▶ CSSHOWER::Splitting_Function_Base::GetXPDF	0.540s				(
▶ std::_Rb_tree_iterator<std::pair<int const, bool>>::operator++	853.381s				(
▶ __ofred_reduction_dp_MOD_fourpoint_reduction_ol	809.926s				(
▶ PDF::LHAPDF_CPP_Interface::GetXPDF	668.065s				(
▶ LHAPDF::Interpolator::interpolateXQ2	582.284s				(
▶ __libm_exp_l9	439.165s				(
▶ LHAPDF::KnotArray1F::iq2below	437.621s				(
▶ LHAPDF::PDF::hasFlavor	432.953s				(
▶ MCATNLO::CF_QCD::CplFac	289.112s				(
▶ MCATNLO::Shower::Reweight	288.336s				(
▶ MODEL::One_Running_AlphaS::AlphaSLam	272.446s				(
▶ LHAPDF::GridPDF::q2Knots	271.472s				(
▶ MODEL::One_Running_AlphaS::Nf	209.858s				(
▶ MCATNLO::CF_QCD::Coupling	204.556s				(

viewing 1 of 1473 selected stack(s)

3.3% (28.667s of 857.578s)

```

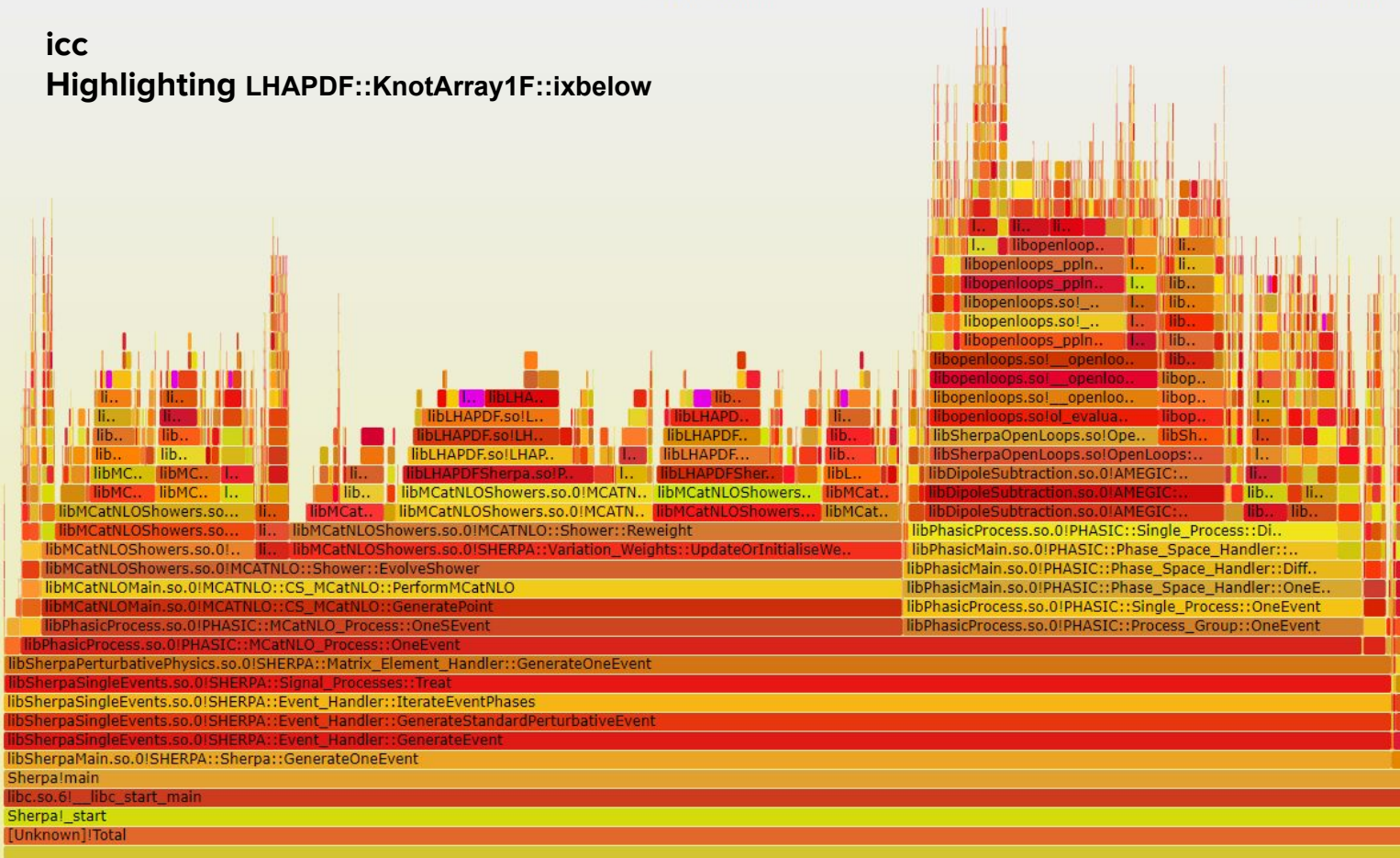
libLHAPDF.so:LHAPDF:(anonymous namespace)::_dxf_dlogx - [unknown source file]
libLHAPDF.so:LHAPDF::LogBicubicInterpolator::_interpolateXQ2+0x25e - [unknown source file]
libLHAPDF.so:LHAPDF::Interpolator::interpolateXQ2+0x116 - [unknown source file]
libLHAPDF.so:LHAPDF::GridPDF::_xfxQ2+0x52 - [unknown source file]
libLHAPDF.so:LHAPDF::PDF::xfxQ2+0x9c - [unknown source file]
libLHAPDFSherpa.so:PDF::LHAPDF_CPP_Interface::GetXPDF+0x35a - LHAPDF::LogBicubicInterpolator::_interpolateXQ2
libMCatNLOShower.so:MCATNLO::Splitting_Function_Base::GetXPDF+0x116 - [unknown source file]
libMCatNLOShower.so:MCATNLO::SF_Lorentz::JFI+0x75 - Splitting_Function_Base::GetXPDF
libMCatNLOShower.so:MCATNLO::Shower::Reweight+0x2d0 - Shower.C3:3
libMCatNLOShower.so:SHERPA::Variation_Weights::UpdateOrInitialiseWeights
libMCatNLOShower.so:MCATNLO::Shower::EvolveShower+0xb5 - Shower.C3:3
libMCatNLOMain.so:MCATNLO::CS_MCATNLO::PerformMCatNLO+0xb3f - ...
libMCatNLOMain.so:MCATNLO::CS_MCATNLO::GeneratePoint+0x421 - CS_MCATNLO::PerformMCatNLO
libPhasicProcess.so:PHASIC::MCatNLO_Process::OneEvent+0x28a - MCatNLO_Process::OneEvent
libPhasicProcess.so:PHASIC::MCatNLO_Process::OneEvent+0x36f - MCatNLO_Process::OneEvent
libSherpaPerturbativePhysics.so:SHERPA::Matrix_Element_Handler::GenerateMatrixElement
libSherpaSingleEvents.so:SHERPA::Signal_Processes::Treat+0x60e - Signal_Processes::Treat
libSherpaSingleEvents.so:SHERPA::Event_Handler::IterateEventPhases+0x3d3 - Event_Handler::IterateEventPhases
libSherpaSingleEvents.so:SHERPA::Event_Handler::GenerateStandardPerturbativeEvent
libSherpaSingleEvents.so:SHERPA::Event_Handler::GenerateEvent+0x590 - Event_Handler::GenerateEvent
libSherpaMain.so:SHERPA::Sherpa::GenerateOneEvent+0x28b - Sherpa.C:22
Sherpa!main+0xa7 - Main.C:27
libc.so.6!__libc_start_main+0xf4 - [unknown source file]
Sherpa!_start+0x28 - [unknown source file]

```



# icc

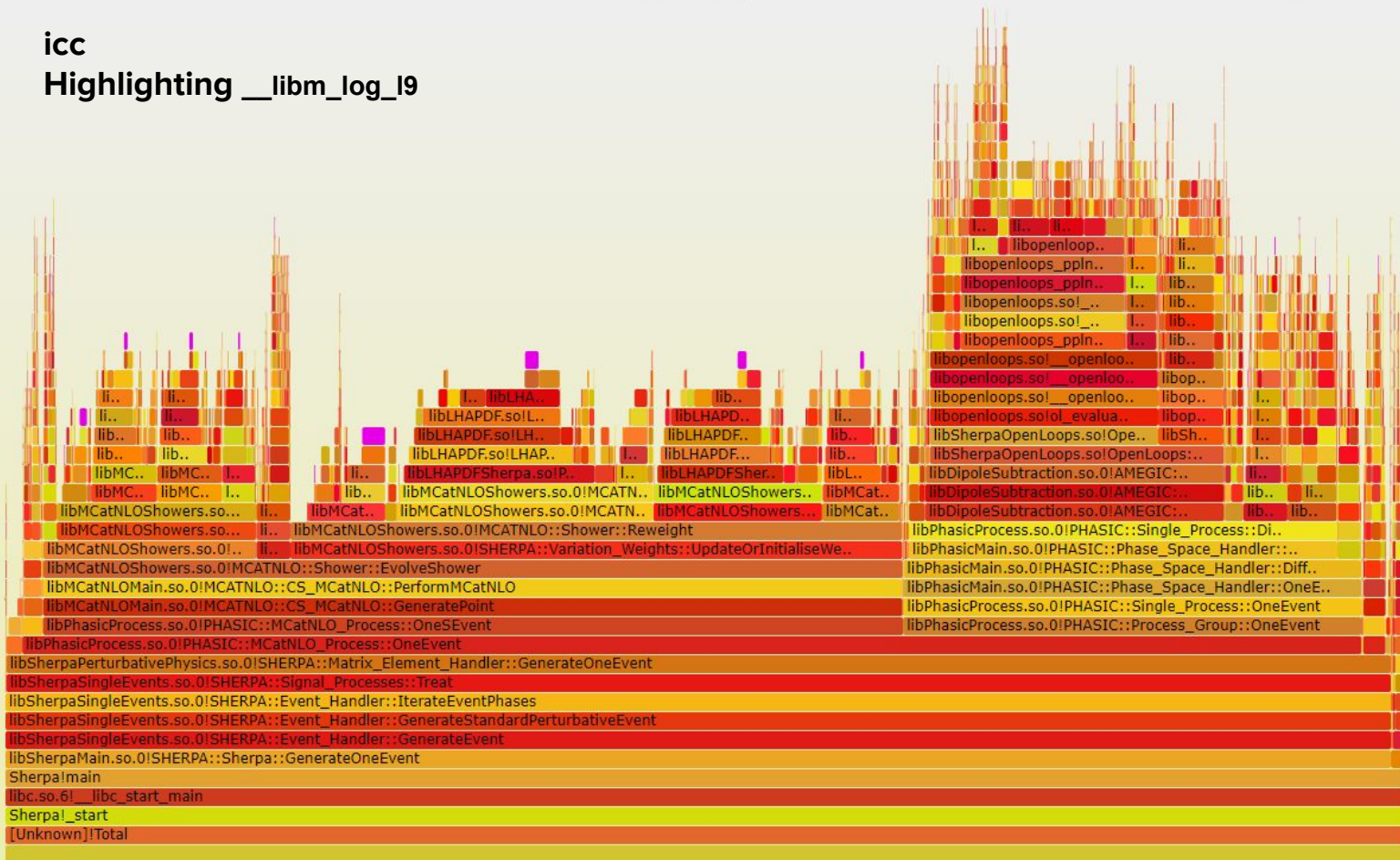
## Highlighting LHAPDF::KnotArray1F::ixbelow





# icc

## Highlighting \_\_libm\_log\_19



# Intel optimisations

- Intel have their own suite of optimisation flags
- Took a little effort to find a combination which would not cause compile errors in either LHAPDF or Sherpa
  - Sherpa: `-xSSE4.2 -O3 -no-prec-div -fp-model fast=2`
  - LHAPDF: `-xSSE4.2 -O3`
- Not currently using link time optimisation (`-flto`), issues...
- Total **18,598 s**
  - **6% faster than GCC**