

Extending testing to include physics regression, a proposal

JIRA Task: [SIM-268](#)



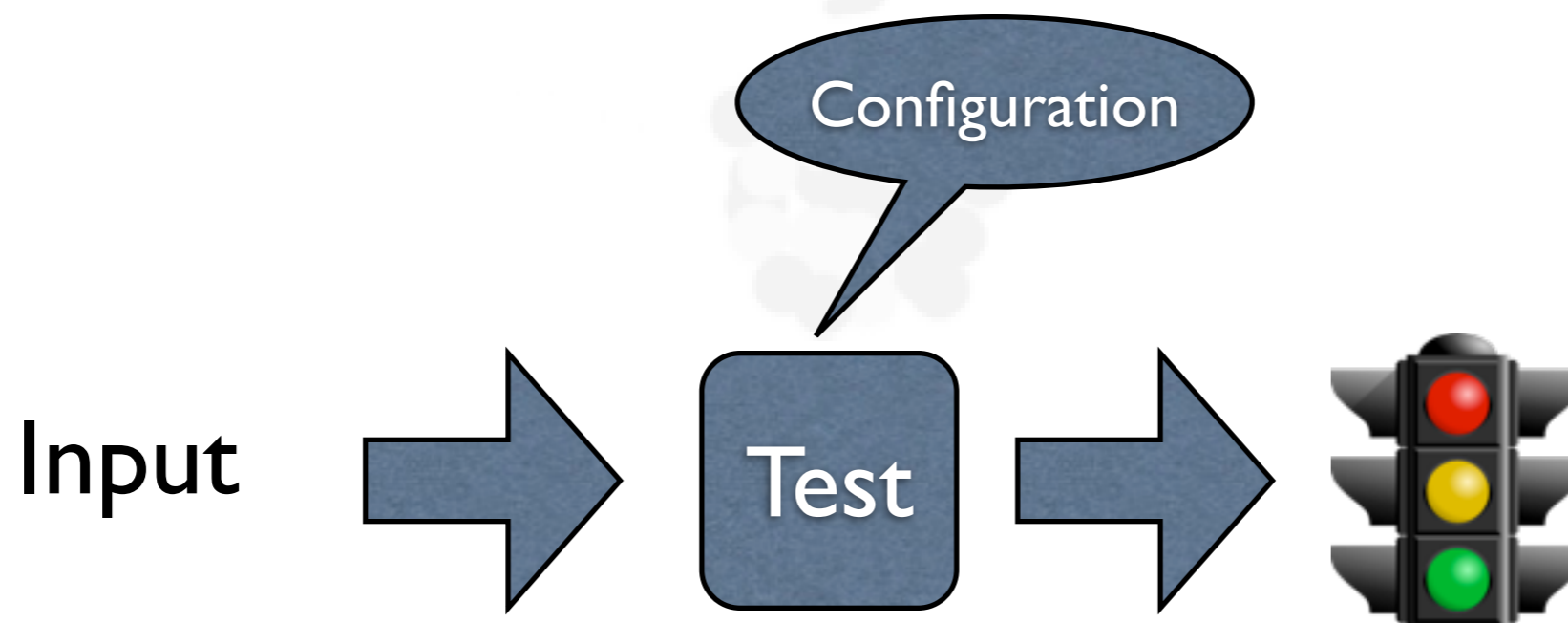
Current Status Analysis

New Requirements

- Current Nightly and CDash system:
 - Code stability oriented
 - Success defined as: Run to completion, stderr empty
- **Investigate automatic extension to include some physics regression testing**
 - Few physics oriented test exists, not unified, interpretation of results and automatization missing
 - Last G4 Week collaboration expressed interest in extending nightlies to include more physics oriented tests
- Geant4 is “unifying” analysis approach: automatic ROOT output
- Requirements:
 - NFR: **Ease to work with**: developers should do minimal work. Ideally in simple cases none
 - NFR: **Ease to interpret results**: a “shift” system could be put in place. Results should be interpreted by non-experts
 - FR: Statistical tests to support: **compare to reference, compare to theory**
 - FR: Both **binned** and **un-binned** distributions
 - FR: System should **provide summary or verbose output**

Basic Idea

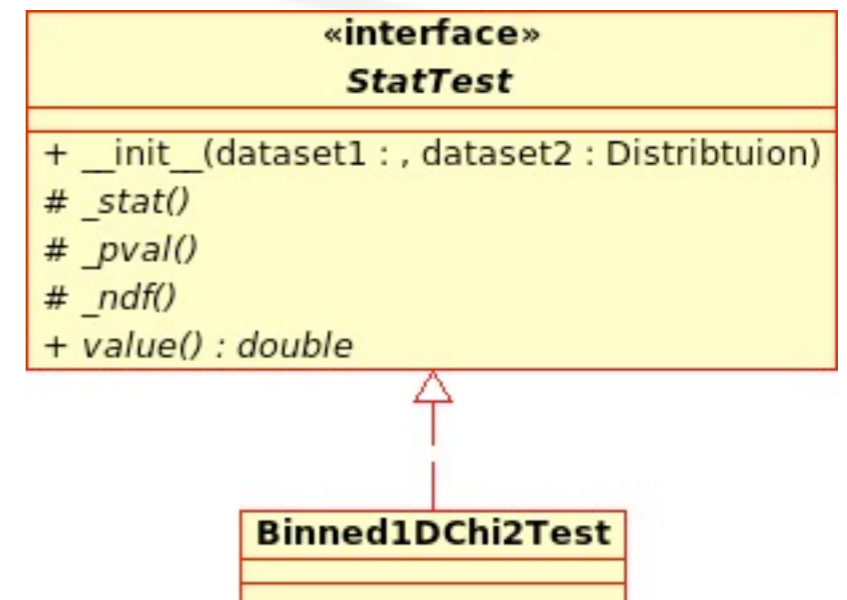
- Derived from my experience in ATLAS' DQ
- Basic idea: an “DQ Algorithm” is a combination of:
 - **An input** (e.g. an Histogram) with simulation results
 - **A Test** (e.g. implementation of Kolmogorov test)
 - **An Output:** a simple flag based on Test output
- A simple way to **configure** define an set of Algorithms is provided
- Pre-requisite: **G4 application should provide ROOT file with TH1 (binned) or TTree (un-binned) distributions**



Implementation

- pyROOT **library of algorithms implemented**
- Use ROOT's implementation of tests
- New tests can be added
- **Simple interface**: each test returns a number: the result (e.g. p-value)

Binned	Un-binned
KS Test	KS Test
Chi2 Test	Anderson-Darling
Chi2 Test (weighted)	



Define Result

- An Output is defined as a **three-state** objects:
 - Two **thresholds** (low, high) are specified and compared to the output of the test R
 - If **$R < \text{low}$** then **FAILED**
 - If **$\text{low} < R < \text{high}$** then **NOTPASSED**
 - Shifter / expert should investigate
 - If **$R > \text{high}$** then **SUCCESS**

Configuration: one file for test

```
Binned = {  
  #Define the defaults  
  'DefaultTestName' : 'BinnedIDChi2Test',  
  'DefaultThresholds' : [ 0.1,0.5],  
  'DefaultReferenceFile' : 'reference.root',  
  
  #Define inputs: histograms a list of dictionaries  
  'Histos' : [  
    #Group 1  
    {  
      'Name' : 'h2',  
      'Thresholds' : [ 0.1 , 0.7 ]  
    },  
  
    #Group 2  
    {  
      'Name' : '/ADirectory/.*',  
    },  
  
  ] #End of Histos defintion  
  
} #End of configuration
```

Defaults

Test number 1: overwrite thresholds

Test number 2: regexp are allowed, test a TDirectory with defaults

Minimal configuration

```
Binned = {  
  'DefaultTestName' : 'BinnedIDChi2Test',  
  'DefaultThresholds' : [ 0.1,0.5],  
  'DefaultReferenceFile' : 'reference.root',  
  
  'Histos' : [  
    #Group 2  
    {  
      'Name' : '.*',  
    },  
  ]  
}
```

Result:

All histograms tested against reference with Chi2 test.
Could be reasonable default for any G4 test

Current Status

- Prototype:
 - Some tests implemented
 - ROOT I/O handling
 - Histograms and TBranches supported
 - Configuration for binned distributions (un-binned missing)
 - **Application that runs tests and produces graphical output (PDF file)**

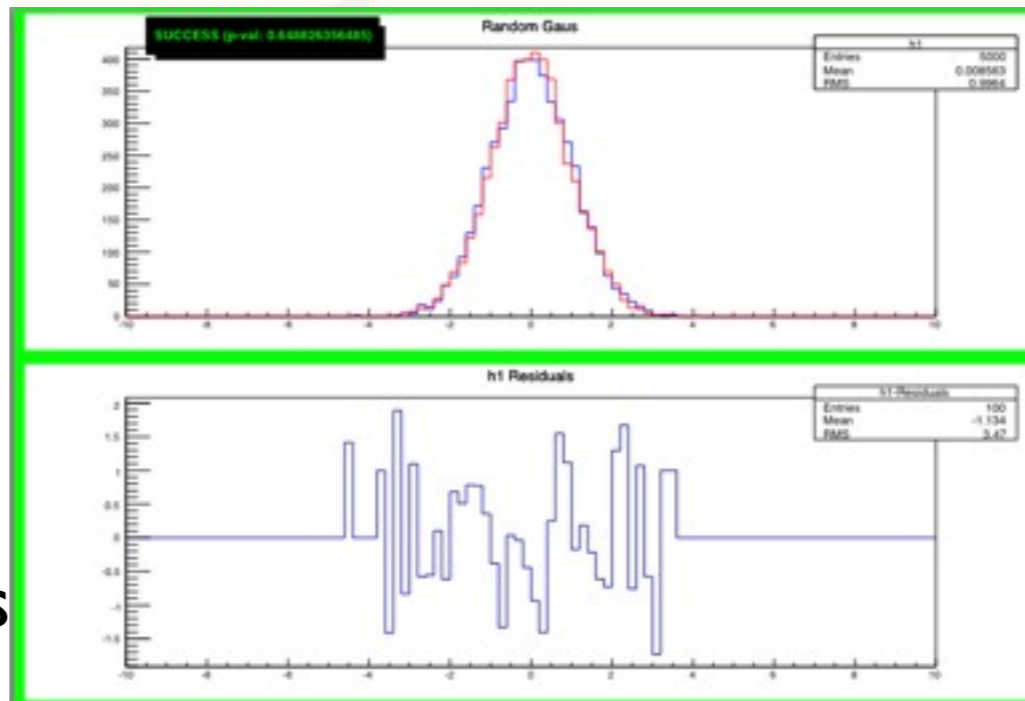
```
$ ./runtests.py testconf.qa test.root
```

```
=====  
h2 : SUCCESS
```

```
/ADirectory/h1 : SUCCESS
```

```
/ADirectory/h3 : SUCCESS  
=====
```

If calculated show residuals



Border: code-color

Conclusions

- Prototype for automatic testing of simulation output
- Each test author should provide:
 - **Reference file**: simply the output of a “good” G4 version
 - **Configuration file**: in simple cases a default can work
- Missing:
 - Conclude un-binned distributions
 - Testing on realistic cases
 - Integrate with CDash
 - Compare to Theory
- If interesting I can propose to collaboration and ask for developer (Validation Task Force)