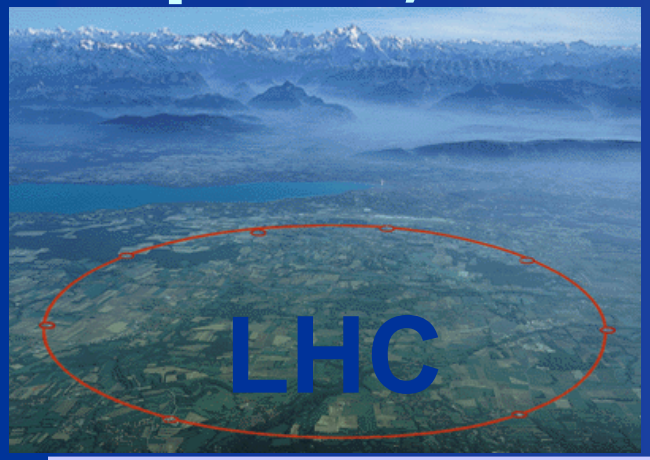
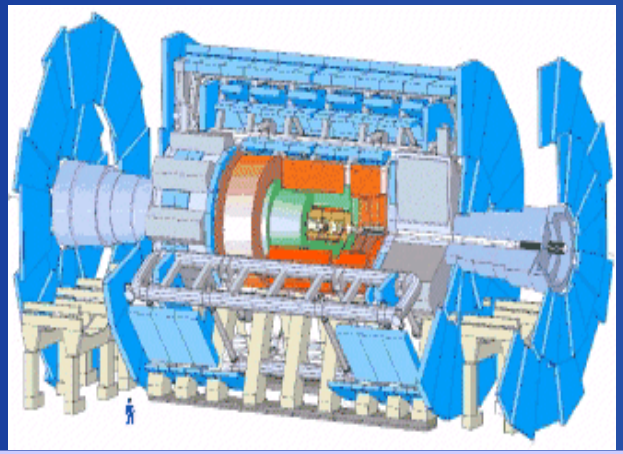
Significance Calculation/A New Method in Searching for New Physics & $H \rightarrow \gamma\gamma$

Yongsheng Gao

Southern Methodist University

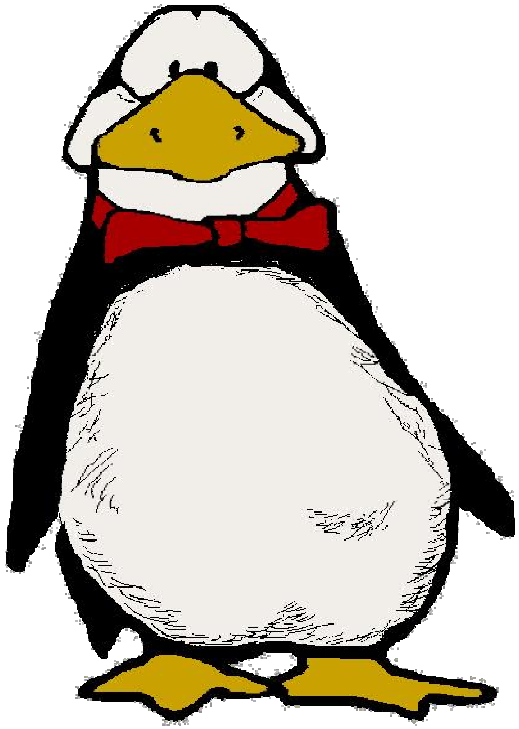
North America SM/Higgs Workshop
April 28, 2006 at Argonne National Lab

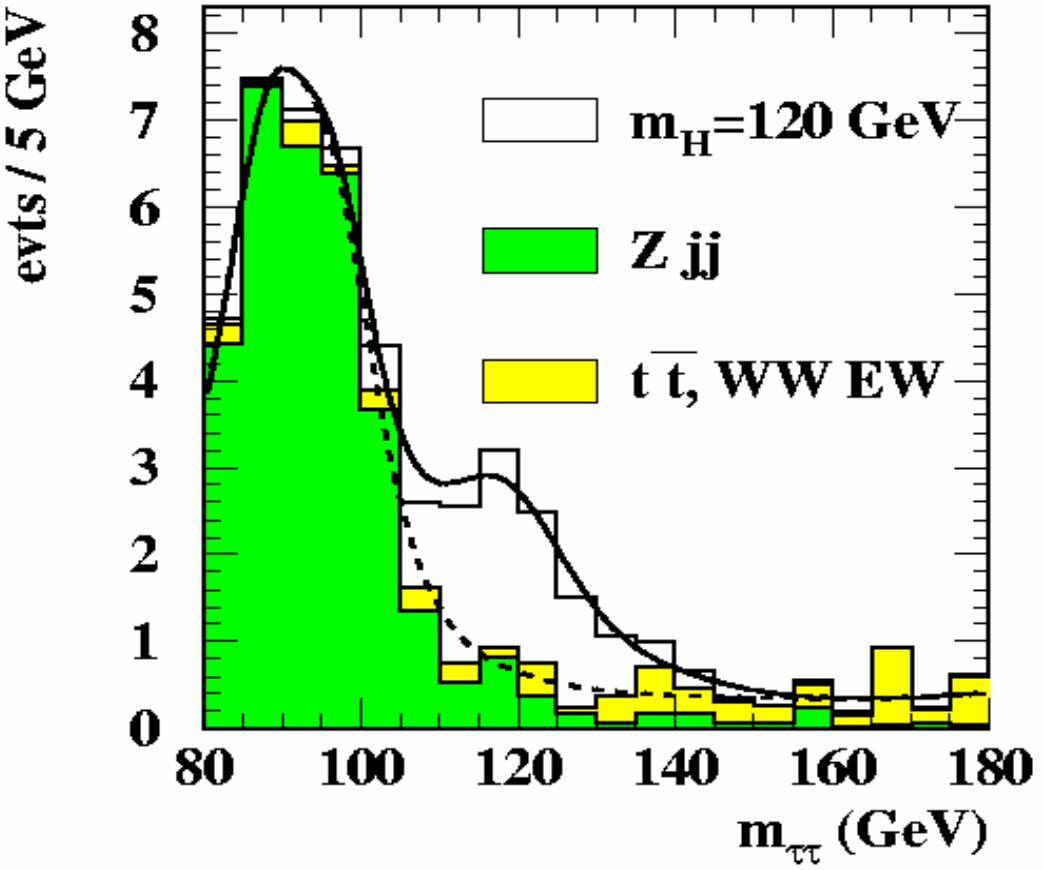


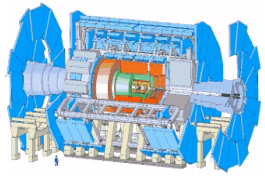


Challenges from first LHC Data

Statistical fluctuation or first sign of New Physics?

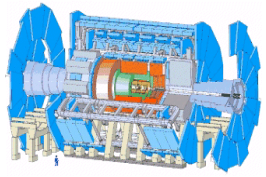






Outline

- Review/problem of current LHC search methods & significance calculation
- General procedures for searching new physics with an unknown location or shape
- A New Method vs Sliding Window
- Results and Conclusion
- Significance of $H \rightarrow \gamma\gamma$ (ATL-COM-PHYS-2006-009 by I. Koletsou, G. Unal, etc)
- Photon Conversion & $H \rightarrow \gamma\gamma$ (David Joffe)



Search for Signal in HEP

Strategy: Search for excess of events in a certain region of a kinematic observable

Observation Probability: $P(n,B) = e^{-B} B^n/n!$

- Gaussian Statistics if B is large (B > 25)
- Observation significance: $S/\sqrt{B}$ where S=n−B

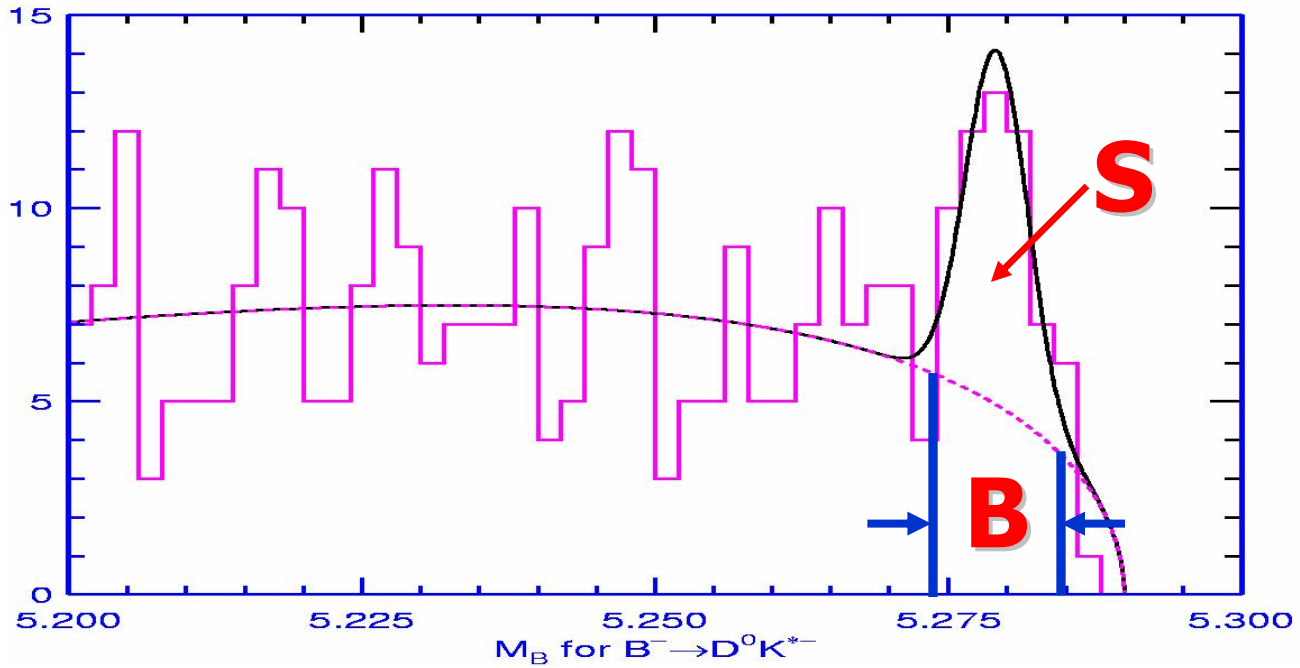
HEP definition of observation significance:

Definition	1σ	2σ	3σ	4σ	5σ
Observation due to BKG fluctuation	15.87%	2.28%	0.14%	0.0032%	2.9×10^{-7}

Signal with known shape/location

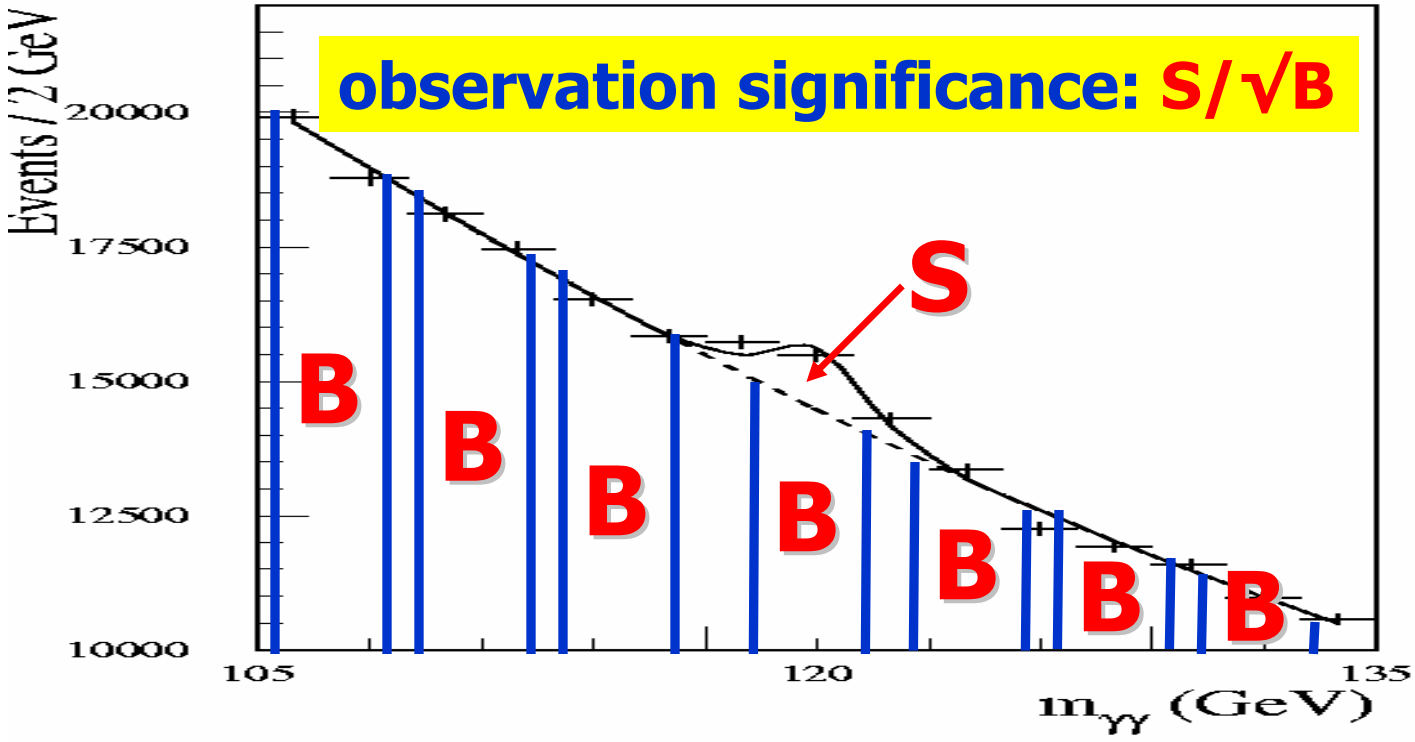
To maximize observation significance S/√B :

- Search region: **±2σ** around **known** Gaussian mean
- Exclude where signal has little chance to show up
- Only valid when **BOTH shape/location** are known



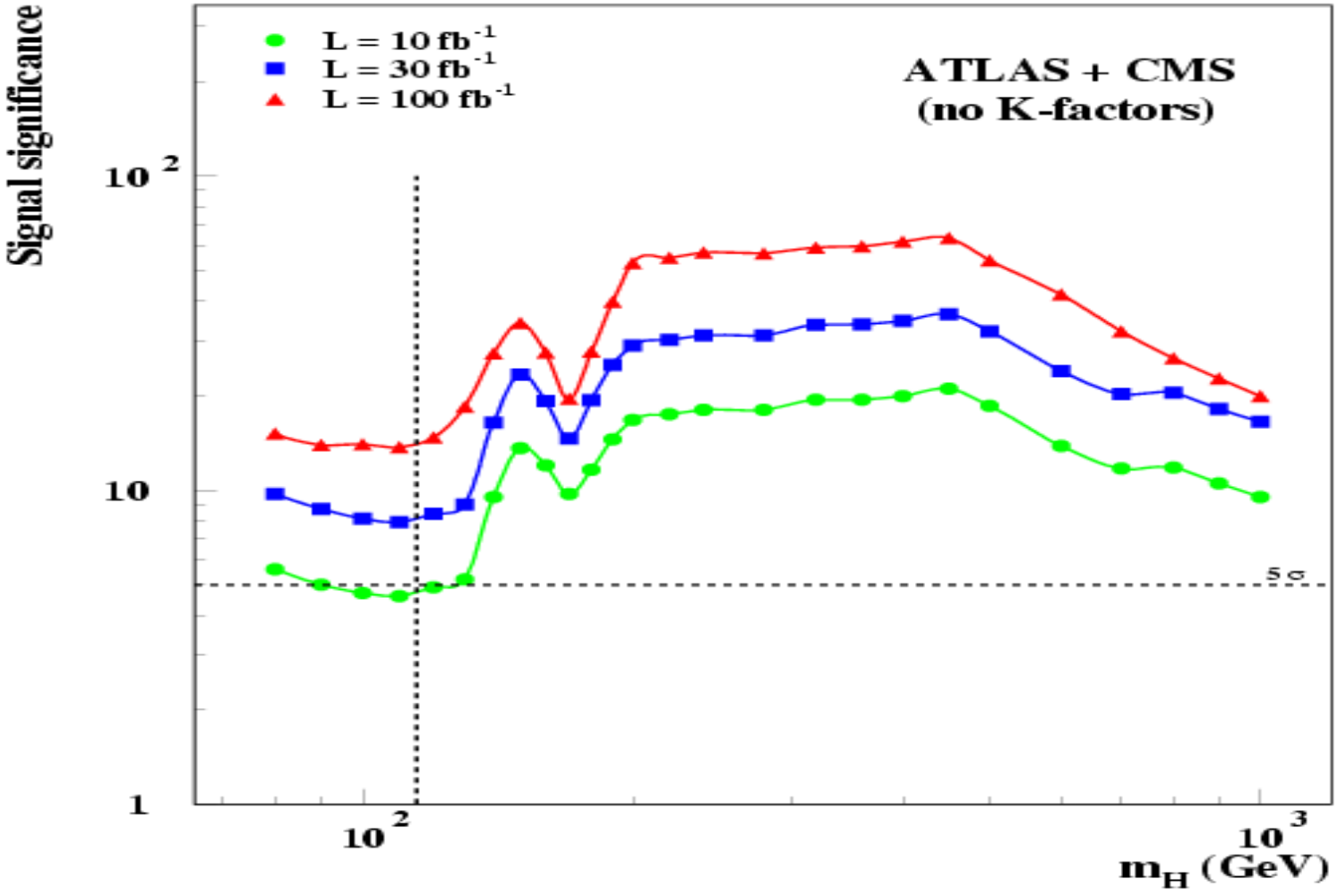
Current LHC Search Method

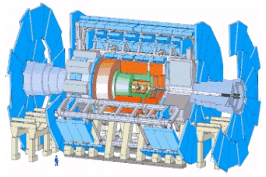
- New Particle's Mass: **Unknown! Range: TeV**
- Use **"Sliding-Window"** and search for excesses
- Significance: Calculated for **each** sliding window



Current LHC Significance Calculation

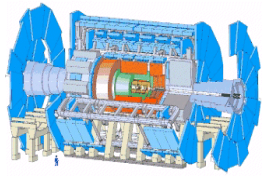
- Each significance (**S/√B**) based on a **~2σ** window





Signal with unknown location

- Explore kinematic region from 0 to 100
- **Background: Flat** distribution from 0 to 100
- **Possible Gaussian Signal:** Stand. Deriv. σ = 1
  Unknown mean from 2 to 98
- "Sliding-Window": **Width=4** and move center from 2 to 98 with **step size of 16,8,4,2,1,0.5,0.2,0.1**
- MC experiment: 500 **background-only** events
- Check how often we find **"significant" (S/√B)** signals from any **background-only** MC exp.

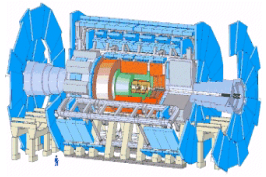


Problem with LHC Search Method

Based on **13,400,000** MC experiments

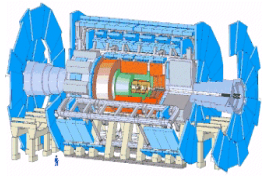
Each experiment contains 500 **background-only** evts

Step-Size	S/√B>1	S/√B>2	S/√B>3	S/√B>4	S/√B>5
16	70.89%	20.42%	1.52%	0.11%	0.002%
8	91.56%	35.25%	2.82%	0.20%	0.003%
4	99.72%	58.53%	5.38%	0.39%	0.007%
2	99.99%	77.86%	9.64%	0.73%	0.015%
1	100.0%	89.03%	14.86%	1.24%	0.027%
0.5	100.0%	94.33%	19.97%	1.83%	0.042%
0.2	100.0%	97.17%	25.42%	2.56%	0.064%
0.1	100.0%	98.01%	28.21%	2.98%	0.078%
Definition	15.87%	2.28%	0.14%	0.003%	2.9×10^{-7}



Problem with LHC Search Method

- "Significance" **over-estimated**
- Multiple tests over **many sub-samples**, but significance is calculated according to **only one sub-sample (False-Positive problem due to multiple-testing)**
- "Significance" depends on **specifics** of the the **search method** (i.e. Step Size) and the **situation** it applies to.



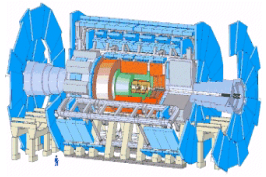
Any Search Method

Type I error rate

- Measures how often **false signals** are claimed when there are only **backgrounds**
- **"Significance" definition in HEP**

Power/Sensitivity

- Measures how often **real signals** can be **correctly identified** when they present



General Procedures

New physics with unknown location/shape:

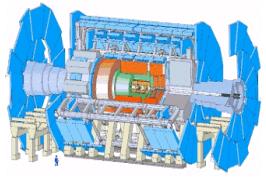
Re-evaluate significance for each method so it follows HEP significance definition using **"Background-only experiments"**

Power/Sensitivity to be determined by **"Signal-Embedded experiments"**

Compare **sensitivity** for different methods

Sensitivity depends on signal location:

- Vary signal location **(Best/Worst scenario)**
- Vary amount of signal embedded



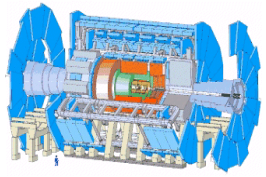
A New Analysis Method

- Apply Maximum Likelihood to **entire** sample
- Find **best fit** to the **entire** sample
  (scan available parameter space)

Re-evaluate significance first (same for SW), so they follows HEP significance definition using **"Background-only experiments"**

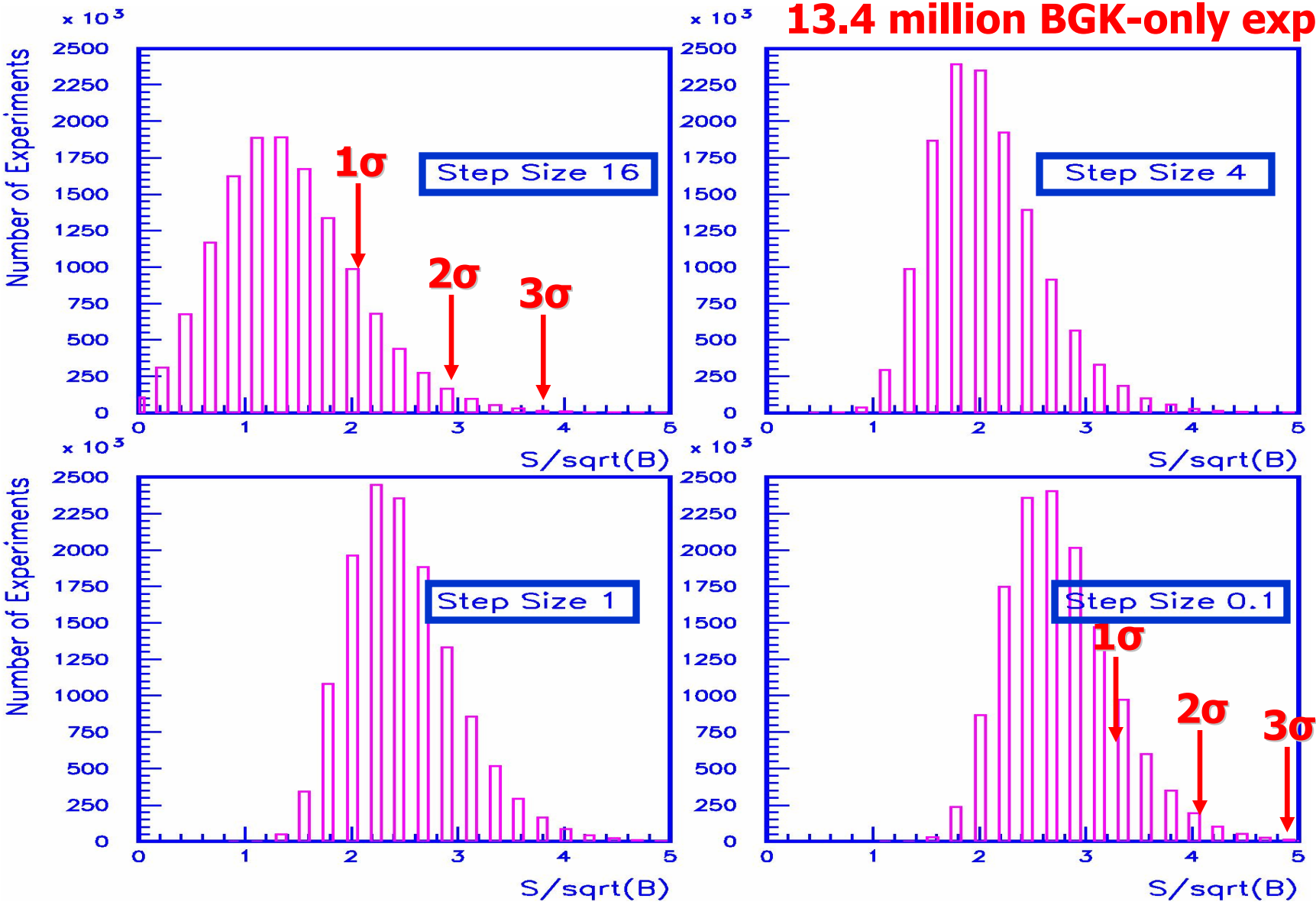
Comparing sensitivities of SW and NM

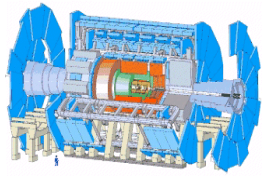
How often signals are found **correctly** with a certain significance (rescaled with "BKG-only")



SW significance recalculation

13.4 million BGK-only exps



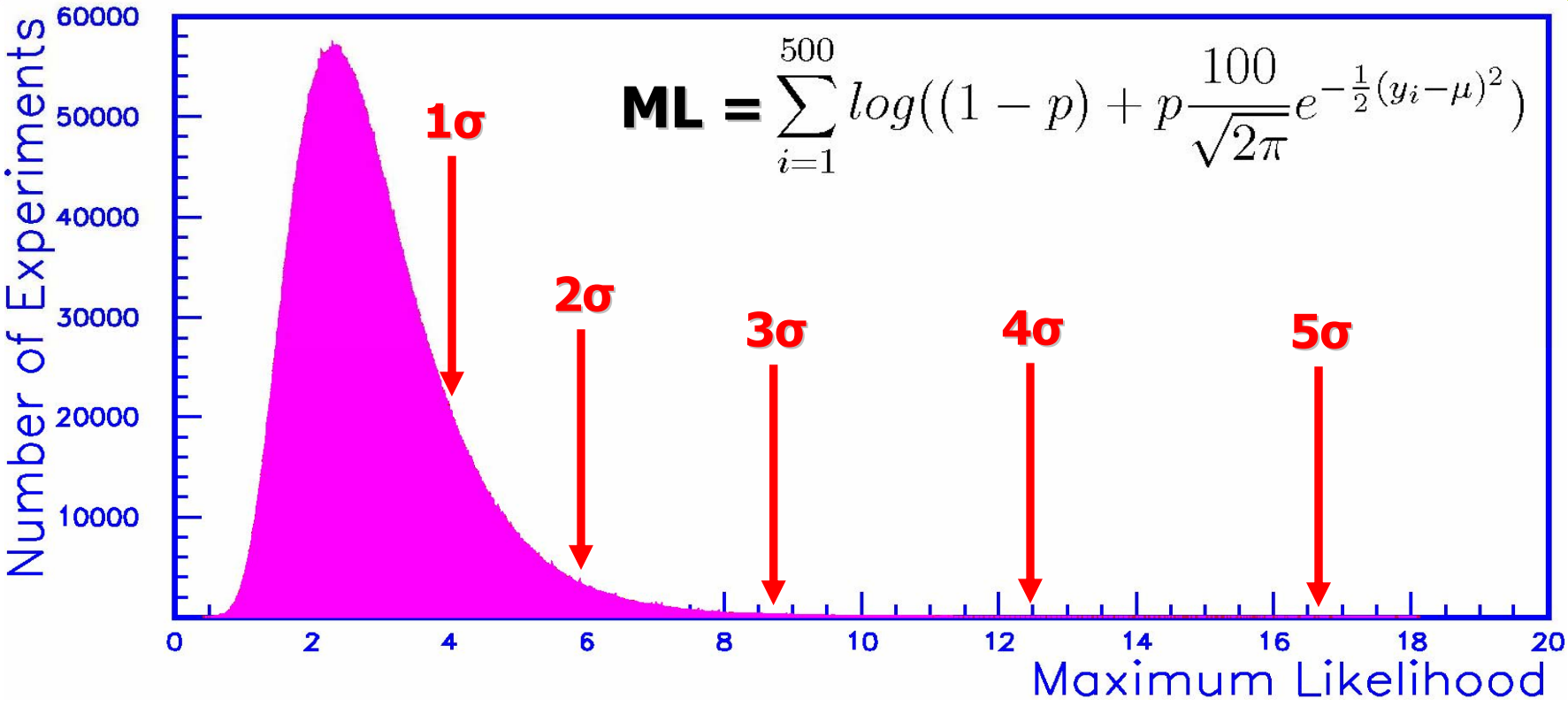


Recalculated significance for SW

S/√B cutoff values using 13.4 million BKG-only exps

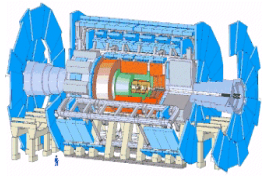
Step-Size	1σ	2σ	3σ	4σ	5σ
16	2.01	2.90	3.80	4.91	6.03
8	2.23	3.13	4.02	5.14	6.26
4	2.68	3.35	4.24	5.36	6.48
2	2.90	3.57	4.47	5.36	6.48
1	2.90	3.80	4.47	5.59	6.48
0.5	3.13	3.80	4.69	5.81	6.93
0.2	3.13	4.02	4.91	5.81	6.93
0.1	3.35	4.02	4.91	5.81	6.93

Recalculated significance for NM



Likelihood cutoff values using 13.4 million BKG-only exps

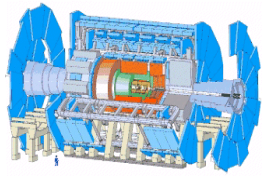
NM	1σ	2σ	3σ	4σ	5σ
Likelihood	4.00	5.94	8.71	12.48	16.61



20 signal embedded, 10K exps

S/√B ~ 20/√20 = 4.47

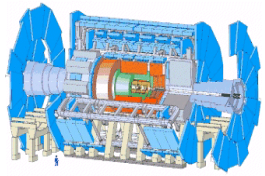
	1σ	2σ	3σ	4σ	5σ
16	9834/0	8969/0	6302/0	2215/0	356/0
8	9694/0	7844/0	5435/0	1616/0	240/0
4	8932/0	7135/0	4542/0	1190/0	158/0
2	8439/0	6867/0	3642/0	1182/0	156/0
1	8086/4955	6365/3913	4032/2509	965/634	174/112
0.5	7795/7264	6812/6273	4020/3782	1046/965	130/100
0.2	7795/7264	6217/5747	3274/3061	1046/965	130/100
0.1	7538/7277	6414/6034	3416/3268	1119/1051	149/111
NW	9116	7406	4073	1209	206



30 signal embedded, 10K exps

$S/\sqrt{B} \sim 30/\sqrt{20} = 6.71$

	1σ	2σ	3σ	4σ	5σ
16	10000/0	10000/0	9963/0	9215/0	5936/0
8	10000/0	9993/0	9917/0	8792/0	5079/0
4	9998/0	9987/0	9841/0	8249/0	4238/0
2	9904/0	9886/0	9633/0	8210/0	4232/0
1	9542/6589	9532/6584	9412/6490	7787/5444	4560/3285
0.5	9249/8784	9246/8781	9152/8665	7743/7331	3847/3683
0.2	9249/8784	9238/8778	9046/8558	7743/7331	3847/3683
0.1	9183/8945	9178/8934	9006/8742	7792/7603	4049/3944
NW	9985	9974	9723	8024	4332

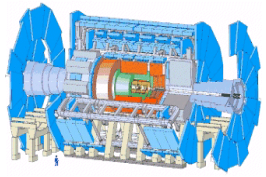


Summary/Conclusion

New physics with unknown location or shape:

Significance calculation non-trivial!

- LHC **"Discovery Potential"** **over-estimated,** depends on **specifics** of the method **(step-size** and the **situation it applies to, etc)**
- **"Sensitivity"** depends on **exact location** of signal, and **specifics** of the method and situation
- **General procedures** for significance calculation and comparison of different search methods
- **A new method** more sensitive to new physics, insensitive to the location of the new physics



Documents/ATLAS work

Talks at Higgs WG in Feb. and Dec. 2005

arXiv:physics/0509174, ATL-COM-PHYS-2005-052
by Y. Gao, L. Lu and X. Wang

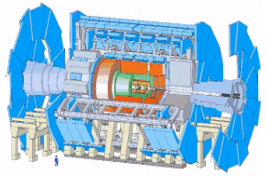
Published in Eur. Phys. J. C45, 659 (2006)

ATL-COM-PHYS-2006-009
by I. Koletsou, G. Unal and L. Fayard

Significance in searching $H \rightarrow \gamma\gamma$ bw 110 to 140GeV

BKG fluctuation prob. increased by a factor of 25!

A real problem to address in ALL searches for New Physics with unknown location or shape!



Significance in $H \rightarrow \gamma\gamma$

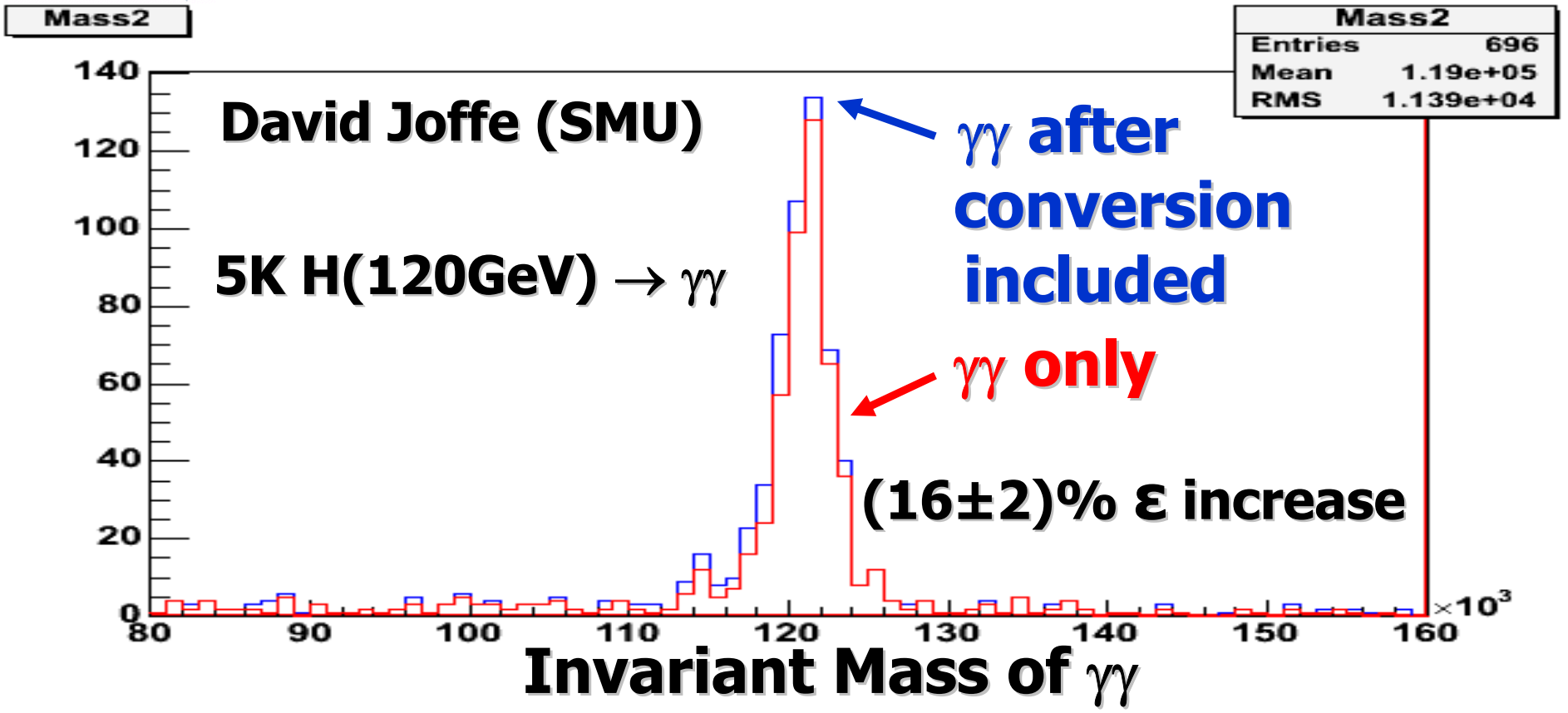
ATL-COM-PHYS-2006-009 (by G. Unal etc):

- Background: Born or Box for 10 fb^{-1}
- $H \rightarrow \gamma\gamma$ Signal: Gaussian
- Method: binned χ^2 fits, also compare with $S/\sqrt{B}$
- "3σ" BKG fluct. Prob.: 0.13% for fixed Higgs mass
- Increased to 3% if mass is bw 110GeV and 140GeV
- Can NOT distinguish a "3σ" signal from a "3σ" statistical fluctuation of background events
- First physics channel to address the significance issue due to unknown location of the signal

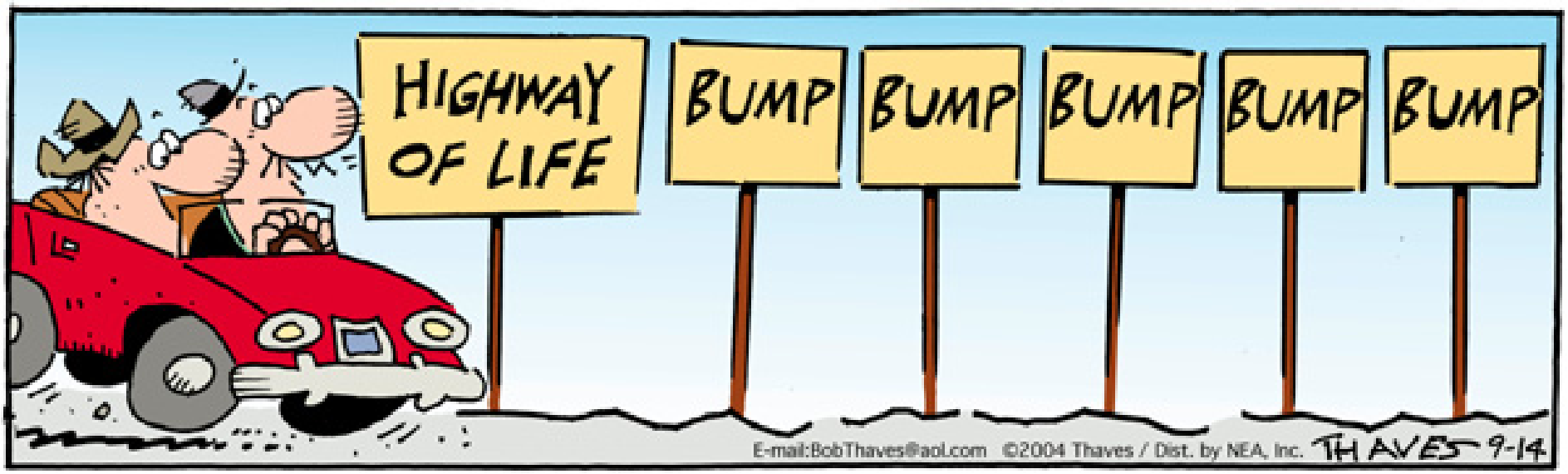
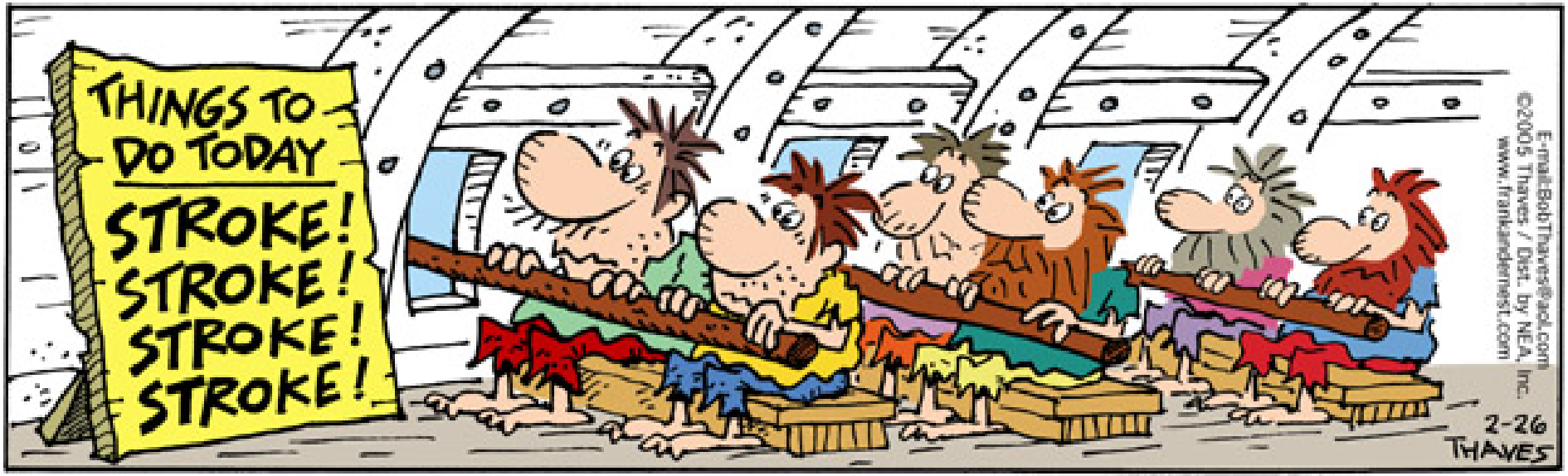
ATLAS/LHC Searches for New Physics:

- Correctly evaluate the significance of each channel
- Maximize discovery potential for each channel

Photon Conversion & H → γγ



David Joffe's talks at e/γ and Higgs WG meetings

F. Yang, B. Mellado, S.L. Wu (Wisconsin)'s talks at Higgs WG meetings

Great discoveries ahead, but