

## ALEXANDER SANDOR SZALAY

Alumni Centennial Professor of Physics and Astronomy Tel: (410)-516-7217  
Department of Physics and Astronomy Fax: (410)-516-5096  
and Professor, Department of Computer Science email: [szalay@jhu.edu](mailto:szalay@jhu.edu)  
The Johns Hopkins University, Baltimore <http://www.sdss.jhu.edu/~szalay>

### Professional Preparation

B.Sc. Physics	Kossuth University, Hungary	1969
M.Sc. Theor. Physics	Eötvös University, Budapest	1972
Ph.D. Astrophysics	Eötvös University, Budapest	1975

### Appointments

Research Associate	Eötvös University, Budapest	1975-1980
Research Associate	University of California, Berkeley	1980-1981
Research Associate	The University of Chicago	1981-1982
Assistant Professor	Eötvös University, Budapest	1982-1986
Staff Scientist	Fermilab	1984-1986
Associate Professor	Eötvös University, Budapest	1986-1987
Professor	Eötvös University, Budapest	1987-present
Professor	The Johns Hopkins University	1989-1998
Alumni Centennial Chair	The Johns Hopkins University	1998-present
Professor of Computer Science	The Johns Hopkins University	2001-present

### Other Academic Experience (six months or more):

Research Associate	University of California, Berkeley	1984
Visiting Professor	The University of Chicago	1984-1986
Visiting Professor	The Johns Hopkins University	1987-1989

### Awards:

First Prize, Storage Challenge, Supercomputing 08	2008
Honorary Doctorate, Eötvös University Budapest	2008
Microsoft Jim Gray Award	2007
Humboldt Research Award, Alexander von Humboldt Foundation	2004
Fellow, American Academy of Arts and Sciences	2003
Eötvös Medal of the Hungarian Physical Society	1997
Associate of the Canadian Inst. for Advanced Research	1991
Szechenyi Prize of the Hungarian Republic	1991
Corresp. Member of the Hungarian Academy of Sciences	1990
E.W. Fullam Prize of the Dudley Observatory	1990
International Physics Olympics, Warsaw, Poland, First Prize	1967

## Short Biographical Sketch

Alexander Szalay

Alexander Szalay is the Alumni Centennial Professor of Astronomy at the Johns Hopkins University. He is also Professor in the Department of Computer Science. He is a cosmologist, working on the statistical measures of the spatial distribution of galaxies and galaxy formation. He was born and educated in Hungary. After graduation he spent postdoctoral periods at UC Berkeley and the University of Chicago, before accepting a faculty position at Johns Hopkins.

He is also a professor at the Department of Computer Science. Half of his 10 graduate students currently are from CS. He is also the founding Director of the Institute for Data Intensive Science at Johns Hopkins, an interdisciplinary institute to tackle cross-cutting challenges in sciences related to the data deluge. His group consists of about 35 people today, focused on various aspects of data intensive computing.

He has written over 500 papers in various scientific journals, covering areas from theoretical cosmology to observational astronomy, spatial statistics and computer science, with over 41,000 citations (Harvard ADS).

In cosmology he coauthored several fundamental papers on structure formation and on the nature of the dark matter in the universe. He has contributed much to the field of theoretical astrophysics and large scale structure. Particular accomplishments include

- Biased galaxy formation in a CDM dominated universe
- Structure formation in a neutrino-dominated universe
- Computing the power spectrum in hot, cold and warm dark matter dominated universes (also defined the terms hot/cold/warm dark matter)
- Various measurements of the large scale galaxy power spectrum

He has developed several novel statistical techniques about optimal estimators for galaxy correlations, power spectra, photometric redshifts for galaxies, optimal coadding of multicolor images, PCA-based spectral classification of galaxies and Bayesian techniques applied to spatial cross-matching of different astronomical catalogs.

He is the architect for the Science Archive of the Sloan Digital Sky Survey, the most used astronomy facility of the world today. He has been collaborating with Jim Gray of Microsoft to design an efficient system to perform data mining on the SDSS Terabyte sized archive, based on innovative spatial indexing techniques. The SDSS Science Archive has attracted an unprecedented amount of users, and is considered to be an example for on-line archives of the future.

He has been leading a grass-roots standardization effort to bring the next generation petascale databases in astronomy to a common basis, so that they will be interoperable – the Virtual Observatory. He was Project Director of the NSF-funded National Virtual Observatory.

He was involved in the GriPhyN and iVDGL projects, creating testbed applications for the Computational Grid. In collaboration with Ian Foster and Ioan Raicu, he explored the concept of Data Diffusion, using thermodynamic principles to incrementally reach better load balancing on distributed data-intensive computations by moving the data stochastically towards an optimal layout on the distributed system. He has collaborated with Bob Grossman on high-speed data analytics for more than a decade, and has been part of the TeraFlow project and the Open Science Grid.

He was also heavily involved in the Data Conservancy, an NSF funded DataNet project, researching the long term curation and preservation of scientific data.

From 2005 he collaborated with Simon White and Gerard Lemson (MPA, Garching) to build a SkyServer-like database out of the Millennium cosmological simulation, which became the reference cosmology simulation used by astronomers all over the world. In collaboration with Piero Madau (UCSC), he is building the 1.2PB database for the Silver River cosmology simulation, currently running at Oak Ridge. He is in the process in generating a 512 grid of cosmological simulations, spanning a cube 8Gpc on the side, to be stored in a publicly accessible 100TB database,

He has also branched out in other scientific areas. Recently, in collaboration with Randal Burns (JHU CS), he has built the 30TB turbulence database, providing immersive access to a large CFD simulation, where users can launch virtual sensors into the simulation that report back their velocity.

With Andreas Terzis and Katalin Szlavecz he has built an end-to-end wireless sensor system for in-situ monitoring environmental parameters, including CO<sub>2</sub>, and measuring the impact of the soil on the global carbon cycles. With sensors around Baltimore, Brazil, Ecuador and the Atacama desert in Chile, the system has more than 200,000 sensor days of data and several hundred million data points.

Recently he published several papers on system research. He led the team that won the Supercomputing Data Challenge in SC-08. He has coauthored several papers with Gordon Bell, arguing how Amdahl's Laws can be used to revisit data-intensive computing architectures from first principles. Applying these ideas, he built a low power system, using Atom processors with extremely good IO performance per unit power (factor of 15 better than a typical rack server). This idea is the subject of a pending patent application with Gordon Bell and others.

He is currently building the Data-Scope, a 6.5PB system with 500Gbytes/s sequential throughput, from a budget of \$1.2M, utilizing a uniquely balanced system built out of hard disks, SSDs and GPUs, for maximal data flow across the system. The project is funded by the NSF.

He has served on many national level committees, at the NRC, NAS, NASA, NSF and the DOE. He has recently served on an NAS panel on Massive Data. For 4 years he has been part of the US CODATA Council. He has also been presenting at the 2012 NITRD meeting, commemorating the 20 years of the Internet. He will also speak at the White House in March 2012 during the announcement of the "Big Data" effort by the Federal Government. He has also presented several times for the National Science Board on Data Management and Data Policies. Currently he is serving on the Advisory Committee for the NSF Computer Science Directorate, and on the NASA Advisory Board to the Administrator for Information Technology.

In 1990 he has been elected to the Hungarian Academy of Sciences as a Corresponding Member. In 2003 he was elected as a Fellow of the American Academy of Arts and Sciences. In 2004 he received one of the Alexander Von Humboldt Research Awards in Physical Sciences. In 2007 he received the Microsoft Jim Gray Award, and in 2008 became Doctor Honoris Clausa of the Eotvos University, Hungary.

## Talks from 1998-2012

Alexander Szalay

1. Invited talk, AAAS meeting, Philadelphia, Feb 1998
2. Colloquium, UMBC Physics Dept, Feb 1998
3. Seminar, Goddard Space Flight Center, Mar 1998
4. Invited talk, Royal Society Discussion Meeting, London, Mar 1998
5. Invited talk, Novartis Foundation workshop, London, Mar 1998
6. Colloquium, Kossuth University, Debrecen, Hungary, Mar 1998
7. Invited Talk: DIMACS meeting on "Astrophysics and Algorithms", Princeton, May 1998
8. Invited Talk, AAS Meeting, San Diego, June 1998
9. Invited Talk, MPA Cosmology Workshop, Munich, Aug 1998
10. Colloquium, Carnegie-Mellon University, Sep 1998
11. Invited talk, Workshop on "Scientific Data Management Problems", Stanford Linear Accelerator Center, Oct 1998
12. Colloquium, University of California, Santa Cruz, Oct 1998
13. Colloquium, Digital Equipment Corporation, MA, Dec 1998
14. Colloquium, The Johns Hopkins University, Jan 1999
15. Invited Talk, 3D Imaging in the Universe, Walnut Creek, CA, Feb 1999
16. Colloquium, Microsoft Research, San Francisco, March 1999
17. Invited Talk, Conference on Photometric Redshifts, Carnegie Observatories, Pasadena, CA, Apr 1999
18. Invited Talk and Panel Member for Digital Sky, Conference on Advances in Digital Libraries, Baltimore, May 1999
19. Invited Talk, Inner Space/Outer Space, Fermilab, May 1999
20. Invited Talk, Instrumentation and Cosmology, Princeton, June 1999
21. Presentation at NASA Headquarters, July 1999
22. Invited presentation at Fermilab, Aug 1999
23. Invited talk, Dunk Island Meeting on Redshift Surveys, Australia, Aug 1999
24. Presentation at NLNR, Sep 1999
25. Invited talk US/EU Workshop on Large databases, Sep 1999
26. Seminar, Institute for Astronomy, Hawaii, Oct 1999
27. Colloquium, Institute for Astronomy, Hawaii, Oct 1999
28. Invited talk, ADASS Conference, Kona, Hawaii, Oct 1999
29. Colloquium, Eotvos University, Budapest, Oct 1999
30. Colloquium, Stewart Observatory, Tucson, Oct 1999
31. Colloquium, Compaq Computer Co, Nashua, NH, Oct 1999
32. Presentation at JHU convocation, New York, Nov 1999
33. Colloquium, Goddard Space Flight Center, Space Sciences, Dec 1999
34. Physics Colloquium, Caltech, Jan 2000
35. Physics Colloquium, University of Kansas, Lawrence, Feb 2000
36. Physics Colloquium, Kansas State University, Manhattan, Feb 2000
37. Invited talk, CITA Cosmology meeting, Banff, Canada, Feb 2000
38. Invited talk, Royal Astronomical Society Meeting, London, Mar 2000
39. Online demo, Intel Computing Continuum Conference, San Francisco, Mar 2000
40. Colloquium, Eotvos University, Budapest, May 2000
41. Keynote speaker, Objectivity Worldview Conference, May 2000
42. Invited Speaker, Virtual Observatories Conference, Caltech, June 2000
43. Invited Speaker, Mining the Sky Conference, Munich, Aug 2000
44. Invited Speaker, IAU Symposium 201, Manchester, Aug 2000
45. Invited Speaker, Grid Workshop, Edinburgh, Sep 2000
46. Invited Speaker, Xth Codata Conference, Bovino, Italy, Oct 2000
47. Panel Member, 17th Codata Conference, Bovino, Italy, Oct 2000
48. Plenary Lecture, ACAT2000 Conference, Fermilab, Oct 2000
49. Colloquium, University of Pittsburgh, Oct 2000
50. Plenary lecture, ADASS X conference, Boston, Nov 2000

51. Physics Colloquium, Hopkins APL, Jan 2001
52. Invited Lecture, ICWS Conference, Tokyo, Feb 2001
53. Invited Lecture, Conference on Stochastic Processes, Wupperthal, Mar 2001
54. Colloquium, Fermilab, Mar 2001
55. Invited talk, Griphyn meeting, Los Angeles, Apr 2001
56. Talk, Aspen Center for Physics, June 2001
57. Invited talk, HEP01 Conference, Budapest, Hungary, July 2001
58. Invited talk, Microsoft Faculty Summit, Redmond WA, July 2001
59. Colloquium, Institute for Astronomy, Hawaii, July 2001
60. Invited talk, Astrostatistics conference, Penn State Univ, July 2001
61. Seminar, Institute for Astronomy, Hawaii, July 2001
62. Particle Physics Seminar, JHU, Oct 2001
63. International Data Grids, Panelist, Supercomputing 2001, Denver, Nov 2001
64. Invited talk, Workshop on Computational Grids, LISHEP2002, Brazil, Feb 2002
65. Astronomy Colloquium, Princeton, Feb 2002
66. Invited talk, AMPATH conference, Chile Apr 2002
67. Keynote Address, SIAM Workshop on Scientific Data Mining, Arlington, VA, Apr 2002
68. Colloquium, NRAO, Socorro, May 2002
69. Invited talk, Hungarian Academy of Sciences, May 2002
70. Presentation, Sigmod Conference, Madison, WI, June 2002
71. Presentation about the SDSS, AAS Meeting, Albuquerque, June 2002
72. Presentation about the NVO, AAS Meeting, Albuquerque, June 2002
73. Invited talk, NVO conference, Garching, June 2002
74. Invited talk, Sky Surveys workshop, Leiden, June 2002
75. Plenary talk, Conference on Hyperbolic Geometry, Budapest, July 2002
76. Presentation, Microsoft, Redmond, July 2002
77. Seminar, Institute for Astronomy, Hawaii, Aug 2002
78. Invited talk on Data analysis, SPIE conference, Hawaii, Aug 2002
79. Invited talk on Data mining, SPIE conference, Hawaii, Aug 2002
80. Invited talk on Web Services, SPIE conference, Hawaii, Aug 2002
81. 2 lectures at the D. Chalonge School of Astroparticle Physics, Palermo, Sep 2002
82. Seminar, Observatoire Strasbourg, Sep 2002
83. Talk, NASA AISRP conference, Mountain View CA, Oct 2002
84. Invited Talk, Workshop on Data Provenance, Chicago, Oct 2002
85. Talk, LSST Software workshop, Baltimore, Oct 2002
86. Invited talk, NSF Workshop on Next Generation Data Mining, Baltimore, Nov 2002
87. Seminar, Dept. Computer Science, Princeton University, Nov 2002
88. Technical colloquium, STScI, Baltimore, Nov 2002
89. Invited talk, From Here to Infinity, Oxford, UK, Dec 2002
90. Seminar, NASA Headquarters, Jan 2003
91. Seminar, Dept. Physics and Astronomy, Jan 2003
92. Invited talk, 'Sustaining Digital Resources' Conference, Washington DC, Feb 2003
93. Colloquium, Harvard University, March 2003
94. Colloquium, Drexel University, March 2003
95. Invited Talk, CIAR cosmology Meeting, Montreal, March 2003
96. Colloquium, University of Virginia, Charlottesville, Apr 2003
97. Seminar, National Center for Biological Information, Bethesda, May 2003
98. Invited talk, NAS meeting on Electronic Publishing, Washington DC, May 2003
99. Presentation, AAS meeting, Session on Dark Energy, Nashville, May 2003
100. Talk, Conference on "Multiwavelength Cosmology", Mykonos, Greece, June 2003
101. 3 lectures at "School of Astroparticle Physics", Lake Balaton, Hungary, June 2003.
102. Public lecture, "School of Astroparticle Physics", Lake Balaton, Hungary, June 2003.
103. Invited talk, "Databases and the Grid" meeting, Edinburgh, UK, July 2003
104. Colloquium, ROyal Observatory of Edinburgh, July 2003
105. Invited talk, IUA Symposium, "Maps of the Cosmos", Sydney, Australia, July 2003
106. Invited talk, "Frontiers in Distributed Information Systems", San Francisco, July 2003

107. Invited talk, Conference on Galaxy Evolution, Zurich Switzerland, Aug 2003
108. Invited Talk, JENAM2003 Conference, Budapest, Hungary, Aug 2003
109. Presentation, NSF Forum on the Computational Grid, NSF, Sept 2003
110. Talk, NASA AISRP meeting, Pittsburgh, Oct 2003
111. Talk, NRC Committee on Analysis of Large Data Sets, Nov 2003
112. Colloquium, Cornell University, Nov 2003
113. Invited talk, CODATA workshop, Lisbon, Dec 2003
114. Seminar on Advanced Data Analysis, Groningen, Feb 2004
115. Nova seminar, Groningen, Feb 2004
116. Nova seminar, Leiden, Feb 2004
117. Invited talk, CIAR meeting, Banff, Canada, Mar 2004
118. Invited Talk, "Understanding the Universe", NSF workshop on Education, Apr 2004
119. Invited talk, Mitchell Conference on Cosmology, TEXAS A & M University, Apr 2003
120. Invited talk, NSF MPS workshop on CyberScience, Apr 2004
121. Invited talk, JHU Workshop on Digital Archiving, Apr 2004
122. Invited talk, LSST meeting, Tucson, May 2004
123. Invited talk, SCIDATA04 Workshop, Microsoft, Redmond, May 2004
124. Invited talk, JHU Workshop on Particle Theory, Baltimore, June 2004
125. Colloquium, Oxford, June 2004
126. Astrophysics seminar, Oxford, June 2004
127. Talk at NVO session, AAS, San Diego, Jan 2005
128. Talk at NIST Workshop on Materials Science, Jan 2005
129. Astronomy colloquium at University of Chicago, Jan 2005
130. Presentation at the NSF AAAC meeting, Feb 2005
131. Invited talk in the session on Grid Computing, AAAS05, Washington DC, Feb 2005
132. Presentation, NSF Site visit for JHU Particle Physics, Feb 2005
133. Physics Colloquium, Cornell University, Feb 2005
134. Invited talk, NAS meeting on Scientific Data Visualization, Washington DC, March 2005
135. Talk, CIAR Annual meeting, Mt. Tremblant, March 2005
136. Joint Astrophysics Colloquium, Munich, March 2005
137. Plenary talk, NAM05, Birmingham, March 2005
138. Colloquium, Microsoft Research, Cambridge, UK, Apr 2005
139. Colloquium, Institute of Astronomy, Cambridge, Apr 2005
140. Seminar, Institute of Astronomy, Cambridge, Apr 2005
141. Talk, MAGPOP Workshop, Budapest, Hungary Apr 2005
142. Keynote Talk, DOE Science Bowl finals, Washington DC, Apr 2005
143. Colloquium, Dept. of Informatik, Technische Universitat, Munich, May 2005
144. Invited talk, Viz05 Conference, Chicago, May 2005
145. Cosmology Seminar, Max Planck Inst fur Astrophysik, Garching, June 2005
146. Invited talk, Ringberg Conference on Surveys, Ringberg Castle, June 2005
147. Discussion leader, Ringberg Conference on Surveys, Ringberg Castle, June 2005
148. Invited Talk, University of Frankfurt, July 2005
149. Seminar on the Virtual Observatory, MPE, Garching, July 2005
150. Invited talk, "Open Questions in Cosmology", Munich, Aug 2005
151. Invited talk, "PhysStat05" conference, Oxford, Sept 2005
152. Invited talk, ADASS05 conference, Madrid, Oct 2005
153. Invited talk, Conference on the Cosmic Microwave Background, Rome, Italy, Oct 2005
154. Session Chair, Sulak Festschrift, Boston, Oct 2005
155. Presentation at AUI Board meeting, Feb 2006
156. Colloquium, NOAO, Tucson, Apr 2006
157. Invited Talk, Cosmology Conference, Texas A&M, Apr 2006
158. Invited Talk, Digital Libraries Conference, Santa Fe, Apr 2006
159. Colloquium, UC Davis, Apr 2006
160. Talk, Scientific Workflows Workshop, NSF, May 2006
161. Presentation, Keck Foundation Workshop, May 2006
162. Colloquium, Freunhofer Inst. for Applied Mathematics, Kaiserslautern, Germany, June 2006

163. Invited talk, Workshop on Turbulence, Chicago, June 2006
164. Invited talk, BES Science Meeting, Baltimore, June 2006
165. Presentation, MS Workshop on Data Preservations, Mountain View, July 2006
166. Invited talk, Microsoft Faculty Summit, Redmond, July 2006
167. Invited talk, Workshop on Nongaussianity in Cosmology, Trieste, July 2006
168. Keynote talk, Microsoft workshop on the World Wide Sensor Web, Semioohma, Aug 2006
169. Presentation, Google-Nature eScience Conference, Mountain View, Aug 2006
170. Invited talk, IAU Symposium on Virtual Observatories, Prague, Aug 2006
171. Keynote talk, NSF Workshop on Oceanographic Observatories, Washington, Sep 2006
172. Colloquium, Space Telescope Science Institute, Oct 2006
173. Keynote talk, Microsoft/JHU conference on eScience, Baltimore, Oct 2006
174. Invited Plenary Talk, Codata Conference, Beijing, Oct 2006
175. Invited Talk, Codata Conference, Beijing, Oct 2006
176. Colloquium, National Science Foundation, Washington DC, Jan 2007
177. Invited Talk, AAS Conference, Seattle, Jan 2007
178. Science Colloquium, NASA Headquarters, Washington DC, Feb 2007
179. Invited Talk, Microsoft TCI Meeting, Redmond, March 2007
180. Invited Talk, Ringberg Workshop on Large Scale Structure, Munich, Apr 2007
181. Keynote Address, NSF/JISC Workshop on Data repositories, Phoenix, Apr 2007
182. Invited Talk, Ringberg Workshop on Large Scale Structure, Munich, Apr 12, 2007
183. Keynote Address, NSF/JISC Workshop on Data repositories, Phoenix, Apr 18, 2007
184. Invited Talk, CIFAR Workshop on Gravitation and Cosmology, Whistler, May 12 2007
185. Colloquium, University of Heidelberg, Heidelberg, May 22, 2007
186. Talk, MAGPOP Workshop on galaxy spectroscopy, Max Planck Institute für Astrophysics, Garching, May 29, 2007
187. Presentation by video link, NRC Committee on Material Science, Irvine, May 30, 2007
188. Invited talk, Caltech eScience Workshop, Pasadena, June 13, 2007
189. Keynote Address, Wiley-Blackwell Seminar on Electronic Publishing, National Press Club, Washington, DC June 15, 2007
190. Invited presentation, NRC Committee on the Future of Radio Astronomy, Charlottesville, VA, June 16, 2007
191. Keynote Address, AustraliAsia eResearch Conference, Brisbane, June 27, 2007
192. Invited Panel Discussion on eScience, NSF Workshop on Computer-Aided Discovery, RPI, Albany, NY, Sep 5, 2007
193. Invited Talk, Workshop on Frontiers of Astronomy, Washington DC, Sep 12, 2007
194. Invited Presentation, Microsoft Workshop on eScience, Brussels, Sep 19, 2007
195. Colloquium, Datalogisk Institute, University of Copenhagen, Sep 20, 2007
196. Presentation, IVOA Interop Workshop, Cambridge, UK, Sep 28, 2007
197. Colloquium, Dept of Astrophysics, Oxford, Oct 10, 2007
198. Invited talk, Royal Astronomical Society Meeting, London, Oct 12, 2007.
199. Presentation, Library Advisory Committee Meeting, JHU, Oct 27, 2007.
200. Presentation, PanStarrs CDR, University of Hawaii, Nov 5, 2007.
201. Presentation, HPC meeting with IBM, JHU, Nov 20, 2007.
202. Colloquium, Dept. of Physics, Rensselaer Polytechnic Institute, Dec 5, 2007.
203. Invited presentation, NAS Committee on Large Data Sets, Washington DC, Dec 10, 2007.
204. Colloquium, Microsoft SQL Server Group, Redmond, Dec 11, 2007.
205. Presentation, Microsoft Research, Redmond, Dec 13, 2007.
206. Invited talk, AGU, Session on Informatics and Data Management, Dec 14, 2007.
207. Invited talk, Meeting of the American Association of Physics Teachers, Baltimore, Jan 22, 2008.
208. Invited talk, AAAS Meeting, Boston, February 17, 2008.
209. Invited talk, Microsoft Research, Redmond, March 7, 2008.
210. Presentation, Trustees Committee on Academic Affairs, JHU, March 9, 2008.
211. Invited talk, NAS meeting on Predicting IED Activities, Washington DC, March 17, 2008
212. Invited talk, Workshop on Data Intensive Science, Yahoo, Sunnyvale, March 26, 2008
213. Invited talk, Workshop on Large Scientific Databases, Asilomar March 31, 2008.
214. Invited talk, Microsoft Research, Redmond, March 7, 2008.

215. Presentation, Trustees Committee on Academic Affairs, JHU, March 9, 2008.
216. Invited talk, NAS meeting on Predicting IED Activities, Washington DC, March 17, 2008
217. Invited talk, Workshop on Data Intensive Science, Yahoo, Sunnyvale, March 26, 2008
218. Invited talk, Workshop on Large Scientific Databases, Asilomar March 31, 2008.
219. Colloquium, Eötvös University, Budapest, May 8, 2008.
220. Invited talk, CNLS conference on Information Science and Technology, Santa Fe, May 15, 2008.
221. Invited presentation, The Gordon and Betty Moore Foundation, San Francisco, May 16, 2008.
222. Invited presentation, Association of Research Libraries meeting, Coral Gables, FL, May 22, 2008.
223. Invited presentation, DB Summit Meeting, Berkeley, May 29-30, 2008.
224. Keynote Presentation, Jim Gray Tribute, Berkeley, May 31, 2008.
225. Presentation, Jim Gray Search and Rescue Meeting, Berkeley, June 1, 2008.
226. Colloquium, Monterey Bay Aquarium Research Institute, June 2, 2008.
227. Astronomy Colloquium, UC Santa Cruz, June 3, 2008.
228. Seminar, NASA Headquarters, Washington DC, June 13, 2008.
229. Keynote, Joint Digital Libraries Conference, Pittsburgh, June 19, 2008.
230. Invited Presentation, Jason Meeting, San Diego, June 26, 2008.
231. Keynote, SSDBM'08 Conference, Hong Kong, July 9, 2008.
232. Invited talk, COSPAR'08 Conference, Montreal, July 15, 2008.
233. Seminar, Inst. Informatik, Technische Universität, Munich, Aug 12, 2008.
234. Invited Talk, JHU-Moore Workshop on Environmental Sensing, JHU Sep 6, 2008.
235. Invited Presentation, Microsoft SQL Server, Redmond, Sep 16, 2008.
236. Keynote Address, I-Science Workshop, University of Leiden, Oct 15, 2008.
237. Invited Talk, Cloud Computing Conference, Chicago, Oct 22, 2008.
238. Presentation, Storage Challenge, SuperComputing 08, Austin, TX, Nov 17, 2008.
239. Invited Talk, Challenges for Great Surveys, Santa Fe, Nov 20, 2008.
240. Presentation, Microsoft eScience Workshop, Indianapolis, Dec 8, 2008.
241. Presentation and Panel Chair, DOE Exascale Computing Workshop, SLAC, Dec 11, 2008.
242. Keynote Address, IEEE eScience Conference, Indianapolis, Dec 12, 2008.
243. Goddard Space Flight Center Colloquium, Greenbelt, Jan 30, 2009
244. Keynote Presentation, Jim Gray Tribute, Berkeley, May 31, 2008.
245. Presentation, Jim Gray Search and Rescue Meeting, Berkeley, June 1, 2008.
246. Colloquium, Monterey Bay Aquarium Research Institute, June 2, 2008.
247. Astronomy Colloquium, UC Santa Cruz, June 3, 2008.
248. Seminar, NASA Headquarters, Washington DC, June 13, 2008.
249. Keynote, Joint Digital Libraries Conference, Pittsburgh, June 19, 2008.
250. Invited Presentation, Jason Meeting, San Diego, June 26, 2008.
251. Keynote, SSDBM'08 Conference, Hong Kong, July 9, 2008.
252. Invited talk, COSPAR'08 Conference, Montreal, July 15, 2008.
253. Seminar, Inst. Informatik, Technische Universität, Munich, Aug 12, 2008.
254. Invited Talk, JHU-Moore Workshop on Environmental Sensing, JHU Sep 6, 2008.
255. Invited Presentation, Microsoft SQL Server, Redmond, Sep 16, 2008.
256. Keynote Address, I-Science Workshop, University of Leiden, Oct 15, 2008.
257. Invited Talk, Cloud Computing Conference, Chicago, Oct 22, 2008.
258. Presentation, Storage Challenge, SuperComputing 08, Austin, TX, Nov 17, 2008.
259. Invited Talk, Challenges for Great Surveys, Santa Fe, Nov 20, 2008.
260. Presentation, Microsoft eScience Workshop, Indianapolis, Dec 8, 2008.
261. Presentation and Panel Chair, DOE Exascale Computing Workshop, SLAC, Dec 11, 2008.
262. Keynote Address, IEEE eScience Conference, Indianapolis, Dec 12, 2008.
263. Goddard Space Flight Center Colloquium, Greenbelt, Jan 30, 2009.
264. Talk at CESA Collaboration Meeting, Univ. of Tennessee, Knoxville, Mar 4, 2009.
265. Talk at IEEE SIGCSE, Chattanooga, TN, Mar 5, 2009.
266. Talk at CIFAR Cosmology and Gravity Meeting, Mt. Tremblant, Mar 7, 2009.
267. Colloquium, LSU CCT, Mar 10, 2009.
268. Colloquium, IIC, Harvard University, Boston, Mar 11, 2009.
269. Invited Talk, Microsoft External research Meeting, Redmond, Mar 31, 2009.
270. Distinguished Lecture, Rice Univ. Houston, Apr 14, 2009.

271. Keynote Address, IPSN'09 Conference, San Francisco, Apr 15, 2009.
272. Colloquium, Nokia Palo Alto Research Center, Palo Alto, Apr 16, 2009.
273. Colloquium, STScI, Baltimore, Apr 22, 2009.
274. Colloquium, Dept. Computer Science, Univ. of Maryland, College Park, Apr 27, 2009.
275. Seminar, Turbulence group, JHU, May 5, 2009.
276. Colloquium, Jet Propulsion Laboratory, Pasadena, May 28, 2009.
277. Seminar, Max Planck Inst fur Plasma Physik, June 21, 2009.
278. Lectures, Eotvos Summer School for Astrophysics, Balatongyorok, July 11, 2009.
279. Colloquium, Los Alamos National Laboratory, Aug 6, 2009.
280. Invited talk, HEC FSIO Conference, Washington DC, Aug 11, 2009.
281. Invited talk, Data Integration Workshop, NAS, Washington DC, Aug 19, 2009.
282. Conference Summary, IDIES workshop, JHU, Aug 26, 2009.
283. Invited Panel Member, Cluster 09 Conference, New Orleans, Sep 1, 2009.
284. Colloquium, NVIDIA Corporation, Santa Clara, CA, Sep 8, 2009.
285. Presentation, Teragrid XD planning meeting, La Jolla, CA, Sep 16, 2009.
286. Key note, Annual Meeting of the German Astronomical Society, Potsdam, Sep 22, 2009.
287. Presentation, Centenary Meeting, Atomki, Debrecen, Sep 24, 2009.
288. Public Lecture, Atomki, Debrecen, Sep 25, 2009.
289. Keynote, HPC in Astronomy, Pune, India, Oct 12, 2009.
290. Presentation, Microsoft eScience Meeting, Pittsburgh, Oct 16, 2009.
291. Presentation, DOE Workshop on Computational HEP, Rockville, MD, Nov 13, 2009.
292. Presentation, Storage Challenge, Supercomputing-09 Conference, Portland, Nov 18, 2009.
293. Presentation, NVIDIA Booth, Supercomputing-09 Conference, Portland, Nov 19, 2009.
294. Presentation, UKRC e-Science Review, Oxford, Dec 11, 2009.
295. Invited Presentation, AstroInformatics, AAS Meeting, Washington DC, Jan 7, 2010.
296. Invited Presentation, NRC Meeting, Washington DC, Jan 29, 2010.
297. Invited Presentation, AstroInformatics, AAS Meeting, Washington DC, Jan 7, 2010.
298. Invited Presentation, NRC Meeting, Washington DC, Jan 29, 2010.
299. Invited talk, CIFAR Meeting, Lake Louise, Canada, Feb 20, 2010.
300. Keynote Address, NESC, Edinburgh Mar 15, 2010.
301. Invited talk, CFD Meeting, Austin TX, Mar 22, 2010.
302. Talk, Dell Computer Co, Austin TX, Mar 23, 2010.
303. Talk, VAO meeting, GSFC, Mar 26, 2010.
304. Keynote address, Salishan Conf. On High Performance Computing, Apr 26, 2010.
305. Invited talk, CITA25 Conference, Toronto, May 15, 2010.
306. Keynote address, Computational Particle Physics, Aussois, France, May 19, 2010.
307. Invited presentation, NAS Astrobiology panel, Washington DC, June 4, 2010.
308. Invited talk, AstroInformatics Conference, Caltech, Pasadena, June 17, 2010.
309. Lectures at the School of Cosmology, NAOC, Beijing, June 2010.
310. Lecture, Chinese Virtual Observatory Group, Beijing, June 22, 2010.
311. Distinguished Lecture, CWI Amsterdam, June 25, 2010.
312. Talk, Cosmology workshop, Eotvos University, Budapest, July 2, 2010.
313. Invited plenary talk, DOE SCIDAC conference, Chattanooga, July 14, 2010.
314. Invited talk, DOE Future of the Field workshop, Snowbird, July 27, 2010.
315. Presentation on Big Data, SciFoo, Google, July 31, 2010.
316. Presentation on Sensor Networks, SciFoo, Google, Aug 1, 2010.
317. Invited talk, Joist Statistical Meeting, Vancouver, Aug 3, 2010.
318. Seminar, Microsoft SQL Server Group, Aliso Viejo, Aug 16, 2010.
319. Discussion leader C3E Conference, Santa Barbara, Aug 17-19, 2010.
320. Keynote, UK e-Science All Hands Meeting, Sep 15, 2010.
321. Keynote, NSF Workshop on Data-Intensive Computing, Seattle, Sep 19, 2010.
322. Physics Colloquium, Univ. Pennsylvania, Sep. 22, 2010.
323. Panel Chair, Workshop on Large Databases, SLAC, Oct 5, 2010.
324. Invited Talk, XLDB Conference, SLAC, Oct 7, 2010.
325. Talk, Microsoft e-Science Workshop, Berkeley, CA, Oct 12, 2010.
326. Astrophysics Colloquium, LBL, Berkeley, Oct 12, 2010.

327. Invited talk, Grand Challenges in Computation, La Jolla, Oct 27, 2010.
328. Capstone address, VisWeek 2010, Salt Lake City, Oct 29, 2010.
329. Invited talk, ADASS 2010 Conference, Boston, Nov 8, 2010.
330. Invited talk, NSF Workshop on Open Scientific Data, Arlington, Nov 12, 2010.
331. Keynote, Workshop on Data Intensive Research, Edinburgh, Nov 22, 2010.
332. Talk, NSF Workshop on High Performance Computing, Arlington, Dec 3, 2010.
333. Invited talk, Challenges in Astrocomputing Conference, San Diego, Dec 16, 2010.
334. Talk, Data Conservancy all-hands meeting, UCLA, Jan 13, 2011.
335. Invited talk, Feynman 50 TedX Conference, Caltech, Jan 14, 2011.
336. Cosmology Seminar, Dept. of Astrophysics, Oxford, Jan 19, 2011.
337. Hintze Lecture, Dept. of Physics, Oxford, Jan 21, 2011.
338. Invited talk, All-Hands Meeting, Mid-Atlantic Crossroads, College Park, Jan 27.
339. Colloquium, Johns Hopkins Advanced Physics Lab, Feb 7, 2011
340. NICT meeting, Washington DC, Mar 7, 2011.
341. Invited talk, JHU Workshop on Statistical Databases, Mar 16, 2011.
342. Invited talk, National Science Board meeting on Data Policies, Washington DC, Mar 28, 2011.
343. Colloquium, Dept. Physics, Univ. Zurich, March 30, 2011.
344. Seminar, Dept. of Astronomy, ETH Zurich, March 31, 2011.
345. Mid-Atlantic Crossroads, All-Hands Meeting, College Park, Apr 7, 2011.
346. Invited talk, Could Computing Challenges, Argonne National Lab, Apr 13, 2011.
347. Keynote, Innovations in Data-Intensive Astronomy, Green Bank, May 5, 2011
348. Invited talk, Data Preservation in High Energy Physics, Fermilab, Batavia, May 16, 2011
349. Physics Colloquium, Stanford University, Palo Alto, May 24 2011
350. Keynote, Statistics with Very Large Data Sets, Baltimore, June 1, 2011
351. Presentation, Maryland Congressional Delegation, JHU, June 2, 2011
352. Closing remarks, Data-Intensive Research Workshop, Edinburgh, June 8, 2011
353. Panel member, Wide Field Surveys Meeting, Baltimore, June 16, 2011
354. Presentation, JHU Physics Advisory Committee meeting., Baltimore, June 21, 2011
355. Presentation, India Biotechnology Roundtable, Rockville, June 27, 2011
356. Invited talk, SKA 2011 Conference, Banff, Canada, July 5, 2011
357. Presentation, SCORE committee, Dept. of Defense, July 15, 2011
358. Invited talk, Microsoft Faculty Summit, Redmond, July 19, 2011
359. Invited talk, Santa Cruz Galaxy Workshop, UC Santa Cruz, Aug 8, 2011
360. DOE Workshop on Computational Challenges in Cosmology, Washington DC, Sep 13, 2011
361. Technical Colloquium, STScI, Baltimore, Sep 13, 2011.
362. Keynote Address, European Grid Forum, Lyon, France, Sep 20, 2011.
363. Invited talk, C3E Conference, Keystone, CO, Sep 27, 2011.
364. Bromery Lecture, JHU EPS, Baltimore, Oct 20, 2011.
365. Colloquium, Canadian Institute of Theoretical Astrophysics, Toronto, Oct 24, 2011.
366. Keynote, Conference on Computational Physics, Gatlinburg, TN, Nov 3, 2011.
367. Colloquium, MIT Lincoln Labs, Boston, Nov 10, 2011.
368. Presentation, NVIDIA Booth, SC-11 Conference, Seattle, Nov 15, 2011.
369. Opening Lecture, HPCDB Conference at SC-11, Seattle, Nov 16, 2011.
370. Panelist, Microsoft e-Science Conference, Stockholm, Dec 6, 2011.
371. Presentation, National Science Board, Arlington, VA, Dec 14, 2011.
372. Invited Talk, AAS Annual Meeting, Jan 9, 2012.
373. Invited talk, NITRD Meeting, Washington DC, Feb 16, 2012.
374. Keynote, ISGC Conference, Taipei, Taiwan, Mar 1, 2012.
375. Panel member, FIMA conference, NYC, Mar 20, 2012.
376. Panel member, OST Big Data meeting, Washington DC, Mar 29, 2012.
377. Panel member, ICDE meeting, Washington DC, Apr 3, 2012.
378. Keynote, Petascale Challenge Meeting, Broad Institute, Boston, Apr 5, 2012.
379. Rosseland Lecture, University of Oslo, Norway, Apr 20, 2012.
380. Colloquium, Goddard Space Flight Center, Beltsville, Apr 25, 2012.
381. Keynote, From Data to Knowledge Conference, Berkeley, May 7, 2012.
382. Keynote, InPar2012 Conference, San Jose CA, May 13, 2012.

383. Physics Colloquium, UC Davis, May 14, 2012.
384. Panel Member, Materials Genome Initiative Meeting, Washington DC, May 15, 2012.
385. Keynote, Microsoft Latin American Faculty Summit, Cancun, May 25, 2012.
386. Keynote, DIDC workshop, Delft, Netherlands, June 19, 2012.
387. Astronomy Colloquium, NAOC Beijing, June 21, 2012.
388. Keynote, XLDB-Asia Conference, June 22, 2012.
389. Invited talk, HPC workshop, Cetraro, Italy, June 25, 2012.
390. Lectures, HIPACC Summer School on AstroInformatics, San Diego, July 9-20 2012.
391. Invited talk, MMDS 2012 Conference, Stanford, July 18, 2012.
392. Seminar, DOE Headquarters, Germantown, Sep 5, 2012.
393. Keynote, Astroinformatics 2012, Redmond, Sep 10, 2012.
394. Keynote, HPEC Meeting, Waltham, MA, Sep 12, 2012.
395. Keynote, LSDMA Symposium, Karlsruhe, Sep 25, 2012.
396. Keynote, Genomics Symposium, JHU, Oct 4, 2012.
397. Panel Member, Atkins Symposium, Ann Arbor, Oct 8, 2012.
398. Lecture, Univ. Michigan, Oct 25, 2012.
399. Keynote, ADASS Conference, Urbana, Nov 5, 2012.
400. Astronomy Colloquium, Univ. Illinois, Urbana, Nov 6, 2012.
401. Tutorial lectures on Big Data, Supercomputing 2012, Salt Lake City, Nov 12, 2012.
402. Invited Talk, Millennium Simulation Workshop, MPA Garching, Dec 19, 2012.

## Professional Activities 1998-2012

Alexander Szalay

1998-2005 SDSS: PI on Science Archive  
1998-2001 Member of NASA Space Science Data System Technical Working Group  
1998-1999 Member of IPAC Users Committee  
1998-1999 Member of Theory, Computation and Data Exploration Panel of the NRC/NAS  
Decatal Survey Committee  
1998-2001 Member of the DEEP collaboration  
1998-2001 President of IAU Commission 47, Cosmology  
1998- Member of the GALEX Consortium  
1998-2005 SDSS: Co-Chair of the Large Scale Structure Working Group  
1999-2002 JHU Application Project Coordinator on Internet II  
1999-2002 JHU Coordinator for the Intel "Technology for 2000" Grant  
1999-2000 Time Allocation Committee, NOAO  
2000-2001 Member of Steering Committee, National Virtual Observatory  
2000 Member of NIH Review Panel on a Molecular Atlas of the Brain  
2001 Member of the PRIME team  
2001 Member of SNAP Review Panel, DOE  
2001 Chair of NASA ADP Review Panel  
2001 Member of the NASA/NSF NVO Science Definition Team  
2001-2005 Project Director, National Virtual Observatory Project  
2001 Advisory Board Member, MONTAGE Project, NASA IPAC  
2001-2005 Member of the Management Council, International Data Grid Laboratory  
2002 Member of NASA ATP Review Panel  
2003 NIH Advisory Committee on Interdisciplinary Research/Technology  
2003-2006 Member of the US CODATA Committee, NAS  
2004 HST Archive planning committee  
2004 Participant, National Science Board Workshop on Data Preservation  
2004 Member of US delegation, OECD Workshop, Washington DC, Apr 2004  
2005 Member of NOAO Long Range Planning Committee  
2005 Member of Organizing Committee, Viz2005 conference, Chicago  
2005 Member of Organizing Committee, Marseille Conference on Galaxy Evolution  
2005 Member of Advisory Group for 2020 Science, Microsoft  
2005 Member of LSST Science Advisory Council  
2005-2008 Member of the Board of the LSST Corporation  
2005 Member of NRC Committee on NASA Science Centers' Oversight  
2006 Member of the GALEX Consortium  
2006 Member of Pan-Starrs collaboration  
2006 Member of NRC Committee on NASA Archive Centers Oversight  
2006 Member of the Oversight Committee for the Cornell Theory Center  
2006 Advisory Committee to the German Federal Government on Supercomputing  
2007 Member of Advisory Committee of the Teragrid  
2007 Member, Cyberinfrastructure and Education Task Force, NSF  
2007- Member, GalaxyZoo Collaboration  
2008 Member, SPACE Collaboration  
2009 Member, UKRC Advisory Panel on e-Science  
2010 Member, NSF Task Force on Data-Intensive Computing  
2010- Member, NASA IT Infrastructure Advisory Committee to the Administrator  
2010-2011 Member, NAS Committee on Massive Data  
2012- Member, NSF CISE Advisory Board

## **Postdocs, grad students, staff members and undergraduates directly supervised**

Alexander Szalay

### ***Undergraduate students***

Eniko Regos, Gabor Toth, Joseph Hollosi, Nolan Li (CS), Sam Carliles (CS), Vivek Haridas (CS), Chris Ralston (CS), Andrew Weisman, Josh Cogan, Daniel Crankshaw, Luis Tueros(CS)

### ***Graduate students***

Istvan Szapudi, Mark SubbaRao, Stephen Landy, Robert Brunner, Kathleen Saavik Ford, Adrian Pope, Jason Goldman, Tamas Budavari, Tanu Malik (CS), Amitabh Chaudhary (CS), Michael Specian, Manuchehr Taghizadeh, Jayant Gupchup (CS), Sam Carliles (CS), Nolan Li (CS), Nuala McCullogh, Bridget Falck, Haijun Tian, Xin Wang, Da Zheng(CS), Matthias Lee (CS), Lin Yang, Dongwei Fan, Laszlo Dobos

### ***Postdoctoral Fellows:***

Andrew Connolly, Tom Broadhurst, Michael Vogeley, Takahiko Matsubara, Bhuvnesh Jain, Martin Kerscher, Roy Gal, Txioxio Benitez, Gyula Szokoly, Rita Kim, Domenico Tocchini-Valentini, Tamas Budavari, Arjun Dey, Lori Lubin, Mark Neyrinck, Ching-Wa Yip, Miguel Aragon Calvo, Dmitry Mishin, Sebastien Heinis, Rafael Santos, Xin Wang, Guillem Lavaux

### ***Staff members:***

Ani Thakar, Peter Kunszt, Jan Vandenberg, Alainna White, Rich Ercolani, Dennis Dinger, Jordan Raddick, George Fekete, Maria Nieto-Santisteban, Wil O'Mullane, Stephen Landy, Sadie Lingham, Gordon Richards, Margie Gier, Ed Givelberg, Richard Wilton, Sue Werner, Deoyani Nandekar,