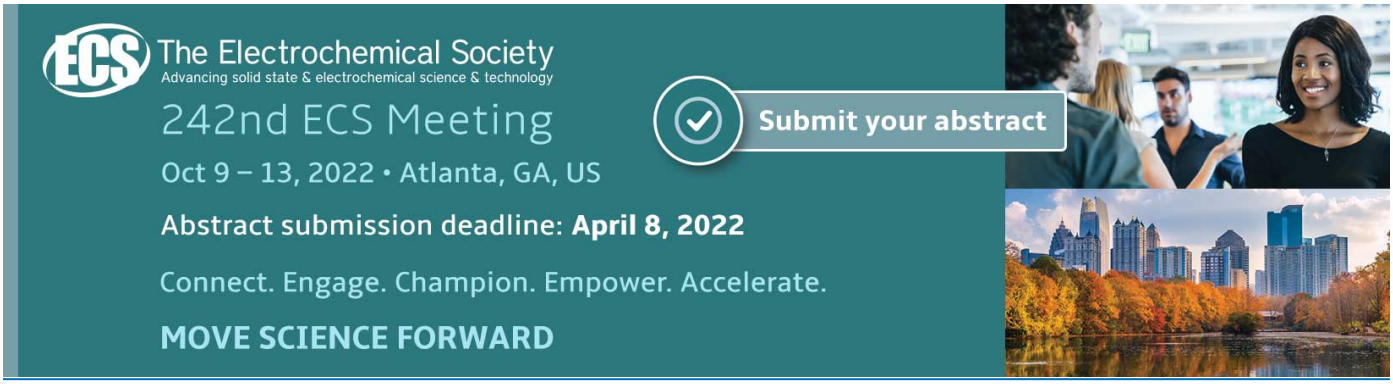
Machine learning for antihydrogen detection at ALPHA



IOP Conf. Series: Journal of Physics: Conf. Series **1085** (2018) 042007 doi:10.1088/1742-6596/1085/4/042007

Machine learning for antihydrogen detection at ALPHA

A Capra for the ALPHA Collaboration
TRIUMF, 4004 Wesbrook Mall, Vancouver BC V6T 2A3, Canada
E-mail: acapra@triumf.ca

Abstract. The ALPHA experiment at CERN is designed to produce and trap antihydrogen to the purpose of making a precise comparison with hydrogen. The basic technique consists of driving an antihydrogen resonance which will cause the antiatom to leave the trap and annihilate. The main background to antihydrogen detection is due to cosmic rays. When an experimental cycle extends for several minutes, while the number of trapped antihydrogen remains fixed, background rejection can become challenging. Machine learning methods have been employed in ALPHA for several years, leading to a dramatic reduction of the background contamination. This allowed ALPHA to perform the first laser spectroscopy experiment on antihydrogen.

1. Introduction

Antihydrogen is a tool to explore the fundamental symmetries [1] that lie at foundation of the Standard Model, i.e., the CPT invariance, and of the General Relativity, i.e., the Equivalence Principle. The CPT symmetry can be tested by comparing the spectra of hydrogen and antihydrogen. The ALPHA collaboration observed for the first time a transition between Zeeman levels of the antihydrogen ground state [2], placed the most stringent upper limit to the antihydrogen electric charge [3], performed the first laser spectroscopy experiment [4] and, more recently, measured the hyperfine splitting of the ground state [5]. An experiment to measure the gravitational acceleration of antihydrogen is currently being built.

In the next section, the ALPHA apparatus is described in some detail with the outline of typical experimental cycle. Successively, the techniques to identify antihydrogen are presented. Finally, the measurement of the 1S-2S transition, depicted in Fig. 1, is described as the prototypical case of using *supervised Machine Learning* for antihydrogen identification. This measurement is a test of CPT invariance with a relative precision of 2×10^{-10}.

2. The ALPHA antihydrogen trapping apparatus

The ALPHA apparatus [6] is located at the CERN Antiproton Decelerator (AD) and its main component are shown in Fig. 2.

Antihydrogen is formed by mixing together a positron and an antiproton plasma. These non-neutral plasma are confined and manipulated in a so-called Penning-Malmberg trap, which consists of an *electrodes* stack and a 1 T solenoid (not shown in the picture).

Cold antihydrogen is trapped near the minimum of the magnetic field gradient produced by the superposition of the field due to the *mirror coils* and to the *octupole*. The depth of

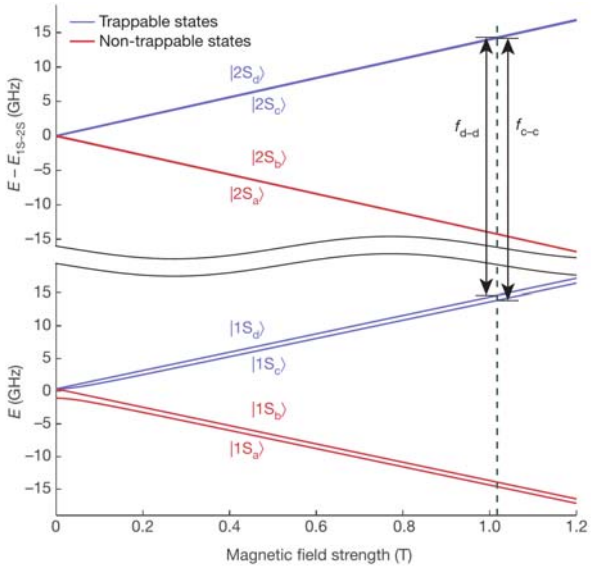


Figure 1. Ground and first excited state of the antihydrogen atom. Since antihydrogen is confined in a magnetic field, Zeeman splitting of the hyperfine levels occur. Only substates where the total spin of the antiatom is anti-parallel to the magnetic field are trappable (blue), the others (red) are not confined in the ALPHA trap. The frequency of radiation is such that it drives a transition between trappable states, as indicated by the black arrows, at a given magnetic field strength.

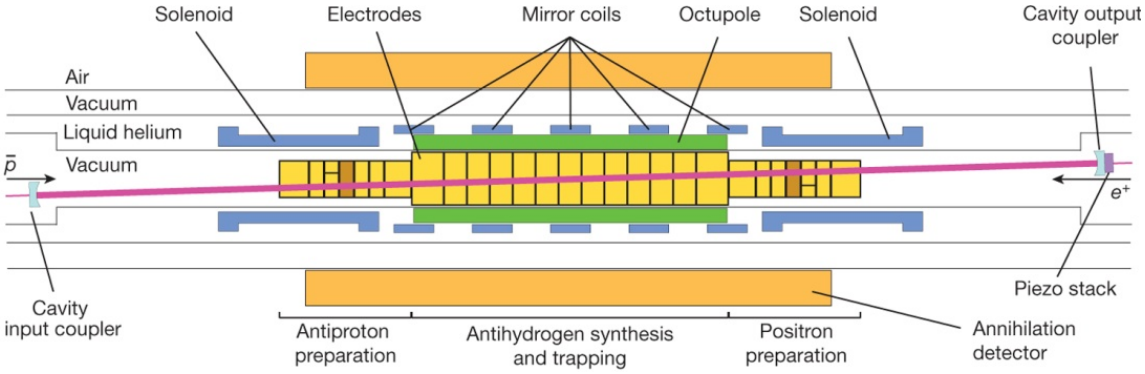


Figure 2. View of the main components of the ALPHA apparatus for the measurement of the 1S-2S transition. The text in italics in the present section refers to the labels in this figure.

the "minimum-B trap" is approximately 50 μeV, or 0.56 K in temperature units, and is largely limited by the maximum current that can circulate in the magnets.

One of the key features of ALPHA is its capability of detecting antihydrogen annihilation. The *annihilation detector* is a silicon strip tracking detector. It is composed of 72 double-sided modules, arranged in three layers. The "silicon detector" permits to monitor the antihydrogen production and to perform the analysis of the response of the antiatoms to the electromagnetic radiation, through the determination - or reconstruction - of the their annihilation location.

The laser system produces 150 mW of 243 nm laser light. In order to increase the light intensity in each direction, allowing the antiatoms to be excited in a reasonable amount of time, the apparatus includes a *power build-up cavity*.

2.1. A typical experimental cycle

In the following the main steps to produce and study antihydrogen are listed with a brief explanation.

Positrons and antiprotons are loaded at centre of the ALPHA apparatus (see Fig. 2), after being cooled by several order of magnitude.

Antihydrogen synthesis in the minimum-B trap. The antiproton plasma is slowly, i.e, on the time scale of 1 s, put into contact with the positron plasma, producing antihydrogen. Most of the antihydrogen created is not magnetically confined, either because has too much kinetic energy or because the bound state is too weak. This phase is referred as "mixing". There are ten expected *cold* antihydrogen atoms confined per mixing. Overall the antihydrogen production takes approximately four minutes and can be repeated several times to trap the desired number of antiatoms.

Trapped antihydrogen is illuminated with electromagnetic radiation, e.g., 243 nm or 121 nm laser light, or 28 GHz microwave, for several minutes. The laser interrogation, for instance, lasts for 600 s.

Trap shutdown where the current in the magnets is ramped down in approximately 1.5 s and all the trapped antiatoms are released from the confinement.

During the mixing phase all the events recorded with the silicon detector are due to the annihilation of "hot" antihydrogen (i.e., antiatoms with too much kinetic energy to be confined) with a negligible background contamination. During the laser (or microwave) irradiation, one expects to see antihydrogen annihilation occurring due to various loss mechanisms. This is the so-called "appearance channel" for antihydrogen spectroscopy measurements. During the first few hundreds of milliseconds of the trap shutdown all the antiatoms, that have not interacted with the electromagnetic radiation, annihilate. This is the "disappearance channel".

The dataset for the experiment under consideration follows a simple protocol, in which three types of trials are conducted: "on resonance", where the build-up cavity is locked to the hydrogen transition frequency, "off resonance", where the laser is detuned by 200 kHz (at 243 nm), and "no laser", where no radiation is present during the 600 s the antiatoms are held in the trap.

3. Detection and identification of antihydrogen

An antihydrogen atom is detected when it annihilates on the trap's wall. On the one hand, the annihilation of positron (a pair of 511 keV gamma rays) is not observed in the current setup. On the other hand, the antiproton annihilation with a nucleon produces mostly charged and neutral π mesons. The π^0 decays rapidly into two high energy gamma rays, which may subsequently produce electron-positron pairs in the surrounding material. The $\pi^{\pm}$ have an average total momentum in the range of 100 - 300 MeV and are tracked in the silicon detector.

Each annihilation event is *reconstructed* through the following - rather standard - steps.

(i) The three-dimensional position of the hits due to the charged particles impinging on the detector is obtained from the charge profile on the strips.

(ii) Tracks are identified by searching through all possible combinations of three hits, one for each of the three layers.

(iii) The identified tracks are fit to helices, since the detector is immersed in an uniform magnetic field.

(iv) The helices are selected so that only the ones most likely due to $\pi^{\pm}$ mesons are used the in the final stage.

(v) The annihilation location, called the *vertex*, is the point where the helices pass closest to one another.

Annihilation events are not the only type of events that trigger the detector readout. The detector occasionally records events due to cosmic rays at rate of $\approx$ 10 Hz. These events constitute the main background to antihydrogen annihilation identification and, as such, must be removed from the data samples. A "cuts analysis" using two variables, as described in [7], achieves a background suppression of 99.9%, which corresponds to a cosmic rays event rate of 47 mHz. This rate is equivalent to 28 background events in the "appearance channel". If one considers that the typical trapping rate is of 10 antiatoms per mixing cycle, the annihilation signal is washed out without a stronger background suppression.

3.1. Background rejection: a classification problem

The difficulty of further suppressing the background arises from events like the ones shown in Fig. 3. The event labelled as "annihilation" (a) was detected during the mixing phase. The cosmic rays events (b,c) were detected when no antiprotons were present in the ALPHA trap. The event (b) is due to a single particle and can be easily classified as a cosmic ray (likely a muon). Even though the event (c) is originates outside the ALPHA apparatus, it is not trivial to classify it as a cosmic ray event.

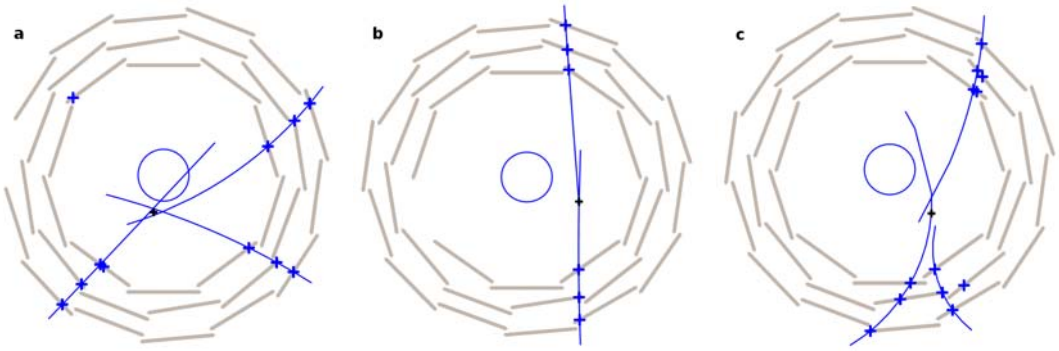


Figure 3. Cross-section view of real events in the ALPHA apparatus. The grey rectangles identify the position of the silicon detector modules. The blue crosses are the hits on the detector and the blue lines represent the reconstructed tracks (helices). a) Antihydrogen annihilation: three tracks event with reconstructed vertex near the trap wall (blue circle). b) Cosmic ray: reconstructed as a "two-tracks" event, even if it is due to single incoming particles. c) Event due to a cosmic ray that produces multiple tracks in the detector and that can be misidentified as an antihydrogen annihilation.

The reduction of the background events to an acceptable rate is only possible by exploiting more sophisticated methods of event classification.

3.2. Training, training samples and discriminating variables

A Machine Learning tool from the SPR package [8] has been employed to classify detected events, achieving high statistical significance in the first spectroscopy experiment on antihydrogen. The classifier of choice is a *Bagged Decision Tree* because of its immediate interpretation.

In ALPHA it is possible to train the classifier on actual data, i.e., without resorting to the generation of a sample of Monte Carlo data. Indeed, the background data sample is collected by operating the silicon detector without antiprotons in the trap, while the signal data sample is derived from the mixing phase, i.e., the annihilation events recorded during the $\overline{\mathrm{H}}$ synthesis. In the present analysis, the background sample consists of $\approx$1.6M events, while $\approx$208k antihydrogen annihilation events compose the signal sample.

It must be stressed at this point that quantities that relate to the antihydrogen annihilation position along the trap axis, i.e., the z coordinate of a Cartesian reference system, cannot be used in the training. This is due to the fact that the antihydrogen annihilation events during the mixing are centred around where the antiproton and positron plasma are confined (see Fig. 1), while the ones due to the interaction with the laser light (or the microwave radiation) depend on the spatial profile of the electromagnetic wave and on the magnitude of the trapping magnetic field.

The training samples are split equally into three subsamples: two for optimization and validation of the training and one for testing, which will be used to evaluate the performance of the classifier.

The following discriminating variables were chosen based on how their distributions for the signal and background event samples can be differentiated.

(i) number of hits N_{hit} (large for annihilation events),
(ii) number of reconstructed tracks N_{helices} (large for annihilation events),
(iii) number of reconstructed tracks used in vertexing $\tilde{N}_{\text{helices}}$ (large for annihilation events),
(iv) ϕ component of the reconstructed vertex (uniform for annihilation events),
(v) r component of the reconstructed vertex (ideally all the annihilation events occur at the trap radius however, due to the finite resolution of the detector, there is a broad peak centred on the expected value, while vertexes due to the cosmic rays do not have a preferred radial position),
(vi) squared residual δ (this is a measure of the tracks straightness, which is greater for the cosmic rays),
(vii) ϕ component of sphericity tensor eigenvector Σ_ϕ (see Eq. (1) and Fig. 4),
(viii) combination of two largest sphericity tensor eigenvalues $\sqrt{\lambda_1^2 + \lambda_2^2}$,
(ix) z component of sphericity tensor eigenvector Σ_z.

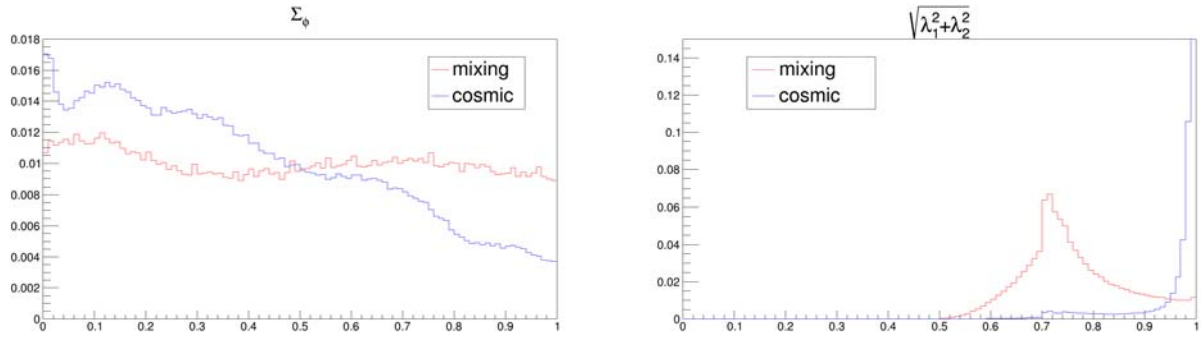


Figure 4. Distributions of the ϕ component (left) and of the combination of two largest eigenvalues (right) of the sphericity tensor, given by Eq. (1) for the mixing and the cosmic rays samples. The distributions are normalized to the total number of events in each sample.

The topology of the detected event can be characterized by the *sphericity tensor*, given by

$$S_{ab} = \frac{1}{N_{\text{helices}}} \sum_{i=1}^{N_{\text{helices}}} \frac{p_a^i \, p_b^i}{|\boldsymbol{p}_i|^2} \,, \tag{1}$$

where $\boldsymbol{p}_i$ is the momentum of the i^{th} reconstructed track and $a, b = x, y, z$. By standard diagonalization of S_{ab}, one finds the eigenvalues λ_j, with $j = 1, 2, 3$, such that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The eigenvector corresponding to λ_1 is labelled $\boldsymbol{\Sigma}$.

3.3. Result of the training for the appearance channel

The classifier training described above is general enough that is adopted for different analysis channels. In this section the results of the training for the appearance channel are presented, since this analysis is the most challenging one, given the long observation time.

The output of the trained classifier[1] on the test sample is shown in Fig. 5. The cut on the classifier output is placed by maximizing the *Punzi significance* [9]

$$\frac{S}{\frac{3}{2}n_\sigma + \sqrt{B}}, \tag{2}$$

where S and B are the number of events classified as annihilation (signal) and background, respectively, and $n_\sigma = 3$ is the number of standard deviations that separate the signal from the background. This yields a background, or "false positive", event rate of 4 mHz ($\pm 7\%$) and the total number of signal events accepted by the classifier is $(37.6 \pm 0.2)\%$ of the mixing sample. The signal acceptance is not too low since the background is suppressed by a factor of twelve and the statistical significance calculated with Eq. (1) is equal to 2.5.

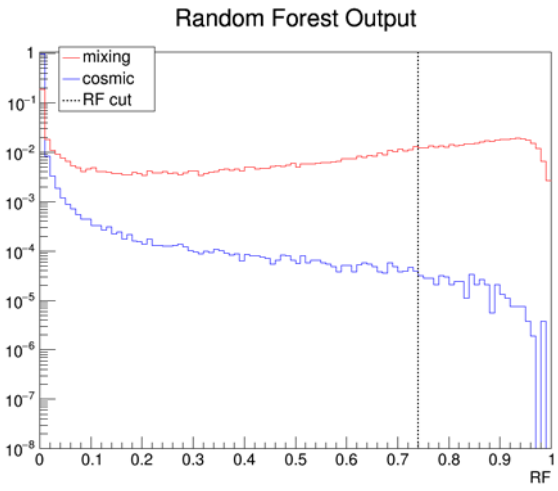


Figure 5. Random Forest classifier (Bagged Decision Tree) with 100 decision tree replicas, 4 variables per tree, 16 events per leaf. The dashed line corresponds to the optimal value of Eq. (2). The distributions are normalized to the total number of events in each sample.

As previously mentioned, it is important to study the z distribution of the reconstructed annihilation vertex. In Fig. 6 it is shown that the z distribution is not distorted by the classifier selection.

4. Laser spectroscopy of trapped antihydrogen

With the trained classifiers at hand, one for the disappearance channel and one for the appearance channel, it is possible to count the antihydrogen annihilation events. A summary of the number of antihydrogen annihilation detected in the two analysis channels is presented in Tab. 1. The on-resonance and off-resonance counts in the disappearance channel differ by 92 ± 15 counts: a clear signature that the resonant laser frequency removed $\approx 58\%$ of the trapped antiatoms. In the appearance channel, the counts in the off-resonance and no laser trials are consistent with the expected number of background events.

[1] The decision trees are grown by optimizing the Gini index.

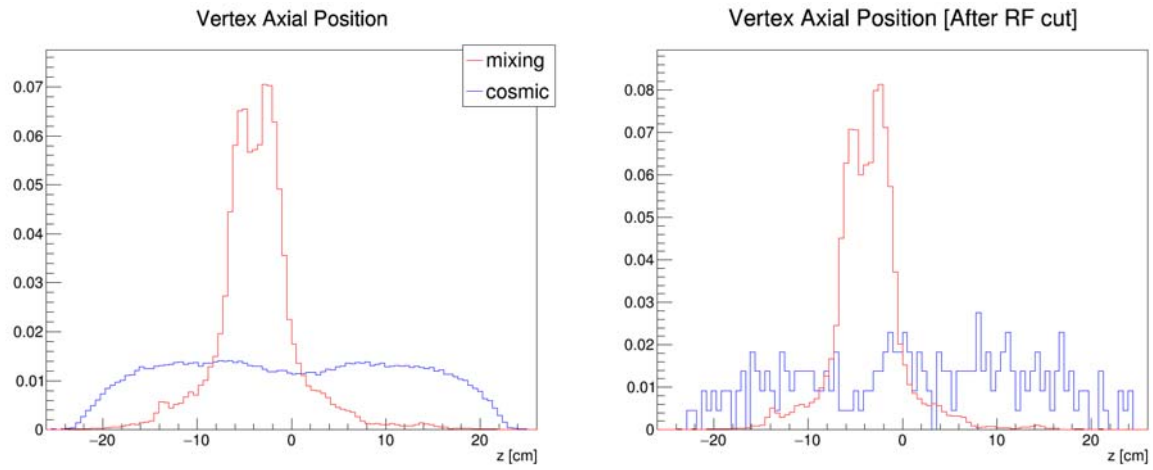


Figure 6. The distributions of the z component of the reconstructed vertex for the mixing and the cosmic rays samples. The distributions are normalized to the total number of events in each sample. The distribution of the mixing sample (left) is not altered after the classifier selection (right).

	Disappearance channel			Appearance channel		
	Background	$\overline{\text{H}}$ events	Uncertainty	Background	$\overline{\text{H}}$ events	Uncertainty
On-resonance	0.7	67	8.2	28.4	79	8.9
Off-resonance	0.7	159	13	28.4	27	5.2
No laser	0.7	142	12	28.4	30	5.5

Table 1. Summary table. The "background" columns indicate the expected number of cosmic rays event. The "$\overline{\text{H}}$ events" columns report the actual number of events classified as antihydrogen annihilation. The "uncertainty" columns are the counting statistics errors.

Taking into account the laser linewidth, the line broadening due to the magnetic field and the detuning of the off-resonance trials, the frequency $\nu_{\text{1S-2S}}$ of the transition is determined to a relative precision of $\delta\nu/\nu_{\text{1S-2S}} \approx 2 \times 10^{-10}$.

5. Conclusion

The adoption of machine learning techniques allowed ALPHA to perform breakthrough experiments on antihydrogen. ALPHA relies upon annihilation position reconstruction for antihydrogen detection. The identification of the antihydrogen annihilation can be challenging due to the cosmic rays background. However, the classifier described here, which is trained on real data, strongly suppressed the false positive event rate, resulting in the first laser spectroscopy experiment on antihydrogen. The analysis presented constitutes a test of CPT invariance at 0.2 ppb level. The current experimental effort consists in determining the lineshape of the 1S-2S transition, where, once again, the use of machine learning classifiers is essential to perform a high precision measurement.

Acknowledgments
The author wishes to acknowledge the whole ALPHA collaboration for the great team effort that lead to this historical measurement: M Ahmadi, B X R Alves, C J Baker, W A Bertsche, E Butler, C Carruth, C L Cesar, M Charlton, S Cohen, R A Collister, S Eriksson, A Evans, N A Evetts, J Fajans, T Friesen, M C Fujiwara, D R Gill, A Gutierrez, J S Hangst, W N Hardy, M E Hayden, C A Isaac, A Ishida, M A Johnson, S A Jones, S Jonsell, L Kurchaninov, N Madsen, M R Mathers, D Maxwell, J T K McKenna, S Menary, J M Michan, T Momose, J J Munich, P Nolan, K Olchanski, A Olin, P Pusa, C Ø Rasmussen, F Robicheaux, R L Sacramento, M Sameed, E Sarid, D M Silveira, S Stracka, C So, G Stutter, T D Tharp, J E Thompson, R I Thompson, D P van der Werf, J S Wurtele. In particular, SS for introducing the machine learning techniques to ALPHA, and JTKM and AO for their tireless analysis efforts.

References

[1] Charlton M, Eades J, Horváth D, Hughes R J and Zimmermann C 1994 *Phys. Rep.* **241** 65–117
[2] Amole C *et al.* (ALPHA Collaboration) 2012 *Nature* **483** 439–443
[3] Ahmadi M *et al.* (ALPHA Collaboration) 2016 *Nature* **529** 373–376
[4] Ahmadi M *et al.* (ALPHA Collaboration) 2016 *Nature* **541** 506–510
[5] Ahmadi M *et al.* (ALPHA Collaboration) 2017 *Nature* **548** 66–69
[6] Amole C *et al.* (ALPHA Collaboration) 2014 *Nucl. Instrum. Meth.* **A735** 319–340
[7] Andresen G B *et al.* (ALPHA Collaboration) 2012 *Nucl. Instrum. Meth.* **A684** 73–81
[8] Narsky I 2005 *arXiv preprint* **arXiv:physics/0507143**
[9] Punzi G 2003 *arXiv preprint* **arXiv:physics/0308063v2**