Shannon entropy of Decay Branching Ratios

Felipe J. Llanes-Estrada*,

Pedro Carrasco Millán, M. Ángeles García Ferrero, Ana Porras, and Esteban Manuel Sánchez García

Dept. Física Teórica I, Univ. Complutense de Madrid, 28040 Madrid, Spain fllanes@fis.ucm.es

How much information is added to the Review of Particle Physics when a new decay branching ratio of a hadron is measured and reported? This is quantifiable by Shannon's information entropy.

We show results at the level of the distribution of decay-channel probabilities, (that is, individual quantum states are integrated over).

We illustrate the concept with some examples.

1) Shannon entropy and its properties

A basic problem in information theory:

How much information is there in a message?

For n bits that take the value 0 with probability p or 1 with probability ($1-p$)

It is a simple combinatorial problem to find *the number of sequences that can be composed*

$$\frac{n!}{(np)!(n(1-np))!} \rightarrow 2^{nS(p)}$$

With Stirling's approximation,

$$\log_2 \frac{n!}{(np)!(n(1-np))!} \approx -p\log_2 p - (1-p)\log_2(1-p) \equiv nS(p)$$

"ENTROPY" of information

Name inspired by the mixing entropy (see box 5)

For a random variable X that takes N values x_i with probability $p(x_i)$, generalize to

$$S(X) = -\sum_{i=1}^{N} p(x_i)\ln p(x_i)$$

*) For an equally distributed variable (box 3) with $p_i=1/N$:

$$S(X)_{max} = -\underbrace{\sum_{i=1}^{N}\frac{1}{N}}_{1=}\left(\ln\frac{1}{N}\right) = \ln N$$

*) S(x) is continuous, positive, convex and additive, that is, if we split a set N into two subsets N_1 and N_2

$$S(N) = pS(N_1) + (1-p)S(N_2)$$

This property makes it useful to analyze particle decays with an increasing number of reconstructed channels

How to choose the base of the log ?

Information in bits: $\log_2$

in "nats": ln

in "hartleys": $\log_{10}$

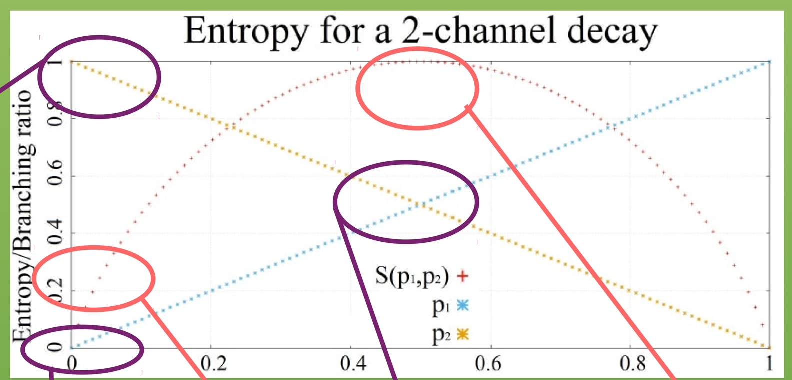
An interesting choice is $\log_N$

In that case the max. S is 1 which allows comparing different "alphabets" (later, the decay of different particles)

A way of seeing S: as the weighted average (with weights $p(x_i)$) of the "information obtained when the variable X takes the value x_i",

$$I_X(x_i) = -\ln p(x_i)$$

3) *Gedanken* cases



Limiting values of the entropy

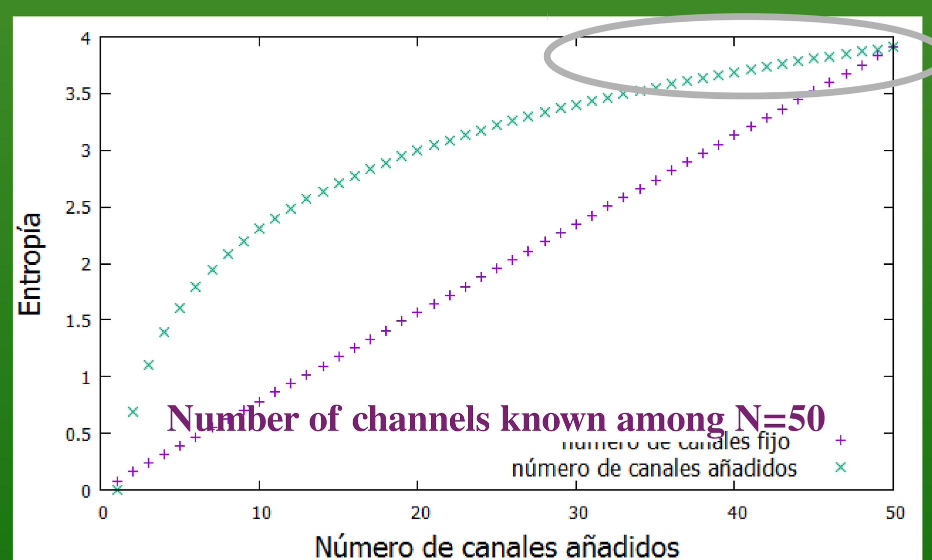
When the probability of the two channels is very different (near 0 and 1, respectively), the entropy is very small

When the two channels are about equally likely, the entropy is maximal; here near 1 because we use bits-base 2 $\log_2 2=1$

This is because the particle almost always decays by the one channel with probability 1. We have maximum information: the decay is predictable.

This is because the particle decays equally by either channel so we cannot tell ahead: minimum predictive power *i.e.* minimum information

Effect of discovering a new decay mode



Upon adding a new channel S grows as a log if all taken with equal branching ratio 1/N

On the contrary, if the additional channels discovered have a decreasingly small branching ratio, then S can be saturated by the first few.

2) Entropy of branching fraction distributions

Partial width to decay through channel f:

$$\Gamma(i\longrightarrow f) = \frac{1}{2m}\int |M(i\longrightarrow f)^2| dLIPS$$

Total width: sum of them all

$$\Gamma_T = \sum_f \Gamma(i\longrightarrow f)$$

Branching ratios

Take BR as the randomly distributed variable belonging to (0,1)

$$BR_{(i\longrightarrow f)} = \frac{\Gamma(i\longrightarrow f)}{\Gamma_T}$$

For each decay channel f

$$S(i) = -\sum_f BR_{(i\longrightarrow f)} log_k BR_{(i\longrightarrow f)}$$

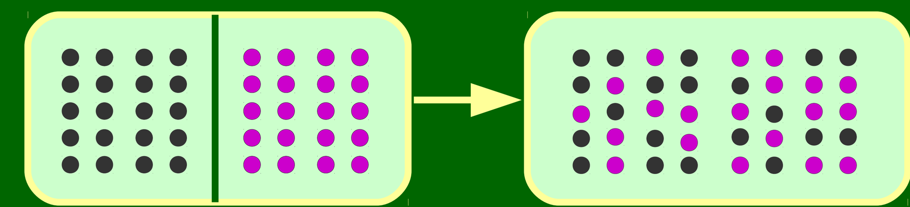
Shannon entropy for the distribution of the branching ratios. All decay channels are assumed to be known. What if some are missing?

$$S(i) = -\sum_{f_{conocidos}} BR_{(i\longrightarrow f)} log_k BR_{(i\longrightarrow f)} - $$
$$-(1-\sum_{f_{conocidos}} BR_{(i\longrightarrow f)}) log_k (1-\sum_{f_{conocidos}} BR_{(i\longrightarrow f)})$$

Group all unknown channels into one with BR given by subtracting all known channels from 1.

5) Where else?

Gibbs Mixing Entropy



With a partition: N_1 N_2

Without a partition: $N_1 + N_2$

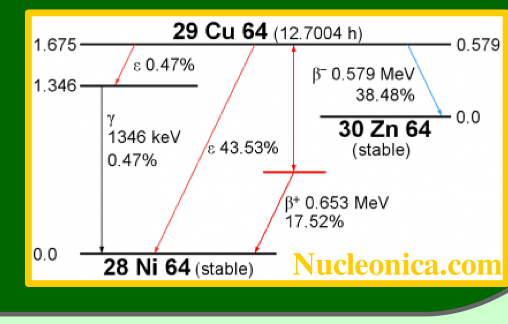
Number of different combinations ("states"): $\Omega = N! / (N_1!\, N_2!)$

Stirling approximation: $\ln N! \simeq N \ln N \longrightarrow \ln\Omega = -N_1\ln(N_1/N) - N_2\ln(N_2/N)$

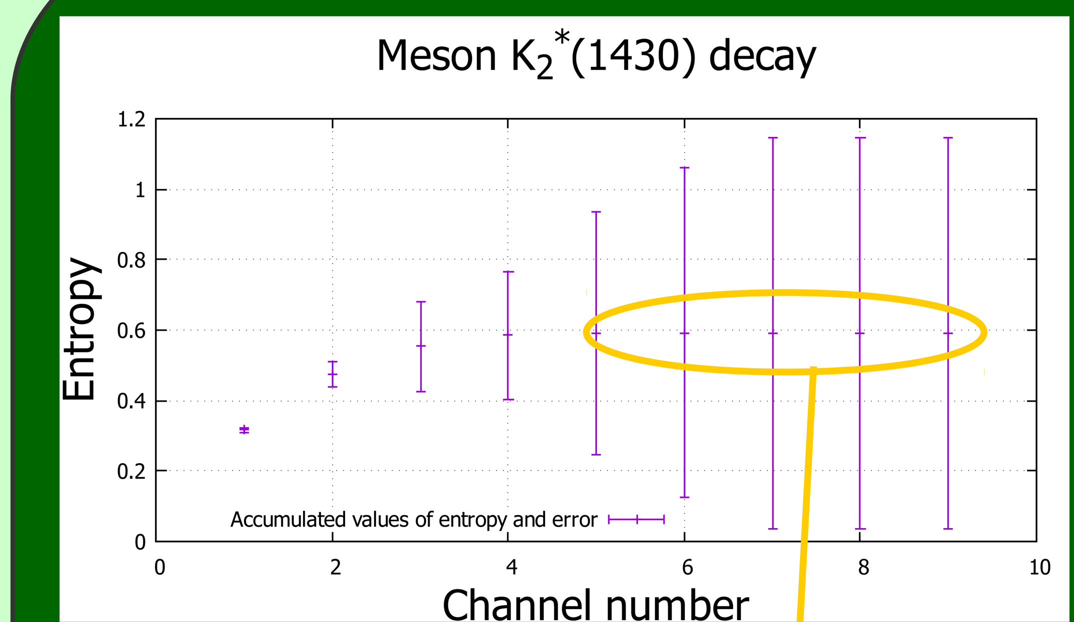
$S = K_B \ln \Omega$; $x_i = N_i / N$ $\longrightarrow S = -NK_B (x_1 \ln x_1 + x_2 \ln x_2)$

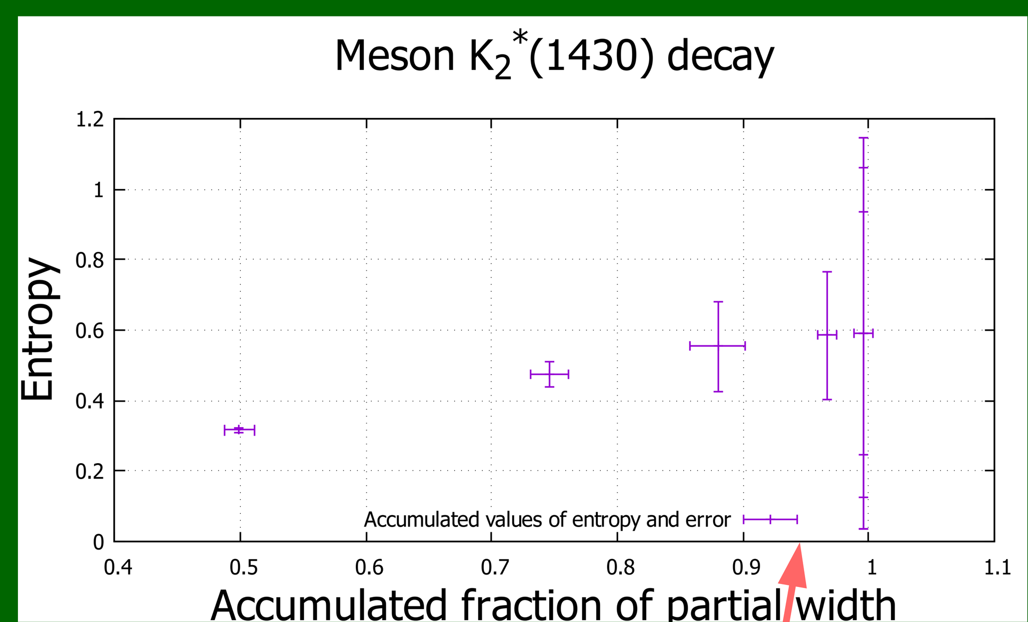
Atomic deexcitation or nuclear decay

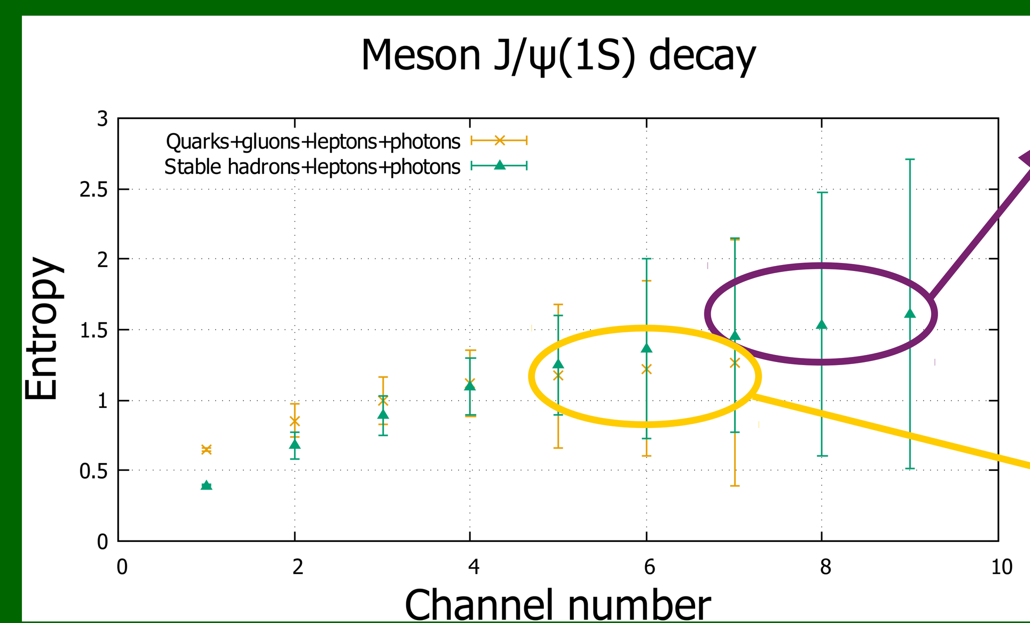
If the decay has several branches the same concepts apply



4)Actual examples from meson decay distributions







While the decays described in terms of hadrons keep accruing entropy S, as they are all small...

In terms of quarks and gluons, the entropy saturates as a few channels account for most of the width Γ.

We add the channels one by one, detracting each time from the generic "$1 - \Sigma\, p_i \ln p_i$"

The entropy saturates when the further channels are of little importance, just the uncertainty (from the PDG BRs keeps increasing)

At that point, most of the width is already accounted for: measuring additional channels barely adds information

The J/y furbishes an interesting example: its decays can be somewhat understood perturbatively in terms of quarks/gluons as an alternative to reconstructed hadrons.