ULB–PhysTh 05/24
November,2005

Introduction to Baryo- and Leptogenesis

J.-M. Frère

Service de Physique Théorique, Université Libre de Bruxelles, CP 228, B-1050 Bruxelles, Belgium

Abstract

Course presented at the ITEP School 2005. These notes aim at an introductory presentation, reviewing in a not-too-technical way the fundamental concepts involved in the baryo/leptogenesis search for the origin of the current excess of matter over antimatter.

Although the title of the course was "leptogenesis" , it starts with reviewing the standard approach through direct baryogenesis, and later explains why leptogenesis is now preferred.

These notes don't aim at being exhaustive, and numerous alternatives to the generation of the baryon number of the universe are not covered.

1 A few concepts

The purpose of this course is to discuss how, from a Universe assumed to be initially symmetrical between matter and antimatter (which can for instance be generated through interaction with gravity), we end up with a Universe clearly domainated by "matter". – Or is it dominated by matter? We can certainly verify it for baryonic matter, and for electrons, but since, as we shall see, the neutrino or antineutrino number is not measured, the total lepton number is unknown. This is why the "matter" vs "antimatter" problem is better known and described as the "origin of the baryon number of the Universe".

At the risk of being pedantic, we will start by a quick review of the origin of these notions.

1.1 Baryon number

The reason this notion was introduced has little to do in fact with the excess of matter over antimatter. The motivation here was nothing less than the stability of the proton;

It is indeed a standard procedure, when an otherwise possible transition is not observed, to introduce a quantum number. The lightest particle carrying such number is then automatically stable if it is assumed that the said quantum number is conserved, or long-lived if the conservation is only slightly violated.

As an illustration of the need for introducing baryon number, it is sufficient to remember that the neutron decay, although it has very little phase space, occurs with an average lifetime of 15 min. , while the lower bound on the proton lifetime (somewhat dependent on the specific decay channel) is of the order of 10^{32} years.

The proton and the neutron were thus (long before the standard model) given baryon number 1, (and -1 for the antiparticles). Assuming all lighter particles to have baryon number 0 makes the proton the lightest particle of its kind, and guarantees its stability to the extent that baryon number is conserved.

This prevents for instance the disintegration $p \to \pi^0 e^+$ which would without this constraint be allowed both from charge and angular momentum conservation (we don't mention lepton number yet here).

1.2 Lepton number

Long considered on a footing similar to baryon number, Lepton number probably does not deserve quite the same status, as the requirements are much less stringent, and there is actually serious reason (beyond the matter-antimatter asymmetry) to consider its possible violation.

Being the lightest charged particle known, the electron is indeed automatically made stable through electric charge conservation alone. So much cannot be said of the μ, or, a fortiori, of the τ leptons, and the latter has many possible decay modes, even taking into account the need of an odd number of spin-1/2 particles in the final state to take into account angular momentum considerations.

As a matter of fact, lepton number and lepton flavour conservation appear more or less at the same level, while baryon number conservation is clearly much a stronger proposition than baryon flavour alone.

Thus, electronic, muonic and "tau" lepton number are introduced, shared each between a charged lepton and its associated "current" neutrino. (we distinguish already between current and mass states.)

In the limit of massless neutrinos, each of these numbers are individually conserved, and so is of course the total leptonic number. In this limit, lepton flavour-violating processes like $\mu \to e\gamma$.

No violation of individual or overall lepton number conservation has been this far observed in charged lepton decays, but solid evidence exists from neutrino oscillations (one neutrino flavour evolves over time into another) that at least individual lepton numbers are violated. The apparent conservation in the charged lepton decays then simply results from the smallness of the neutrino masses compared to the energy scale of the decays considered.

The question of total lepton number conservation stays open, and evidence is most likely to come from low-energy processes, like the neutrinoless double beta decays.

1.3 Evolution of the "fermion number" notion

From a purely phenomenological (ad-hoc) concept, the notion of fermion number has considerably evolved, both on the experimental and theoretical fronts.

First of all, in the context of field theories, like the Standard model, conservation laws are generally associated to invariances of the Lagrangian over continuous (mostly phase) transformations, through the Noether theorem.

For baryons, the formulation now takes place in terms of quarks rather than the baryons themselves (proton, neutron, lambda...). Both the "up" quarks (u,c,t) and the "down" quarks are assigned baryon number 1/3, while their antiparticles have -1/3. The individual numbers which could be associated to various species (like the strangeness) are known to be broken by the mass terms, and thus only the overall baryon number is protected. It should be noted that, despite the fact that all quarks are charged, interactions violating total baryon number are not excluded. They can (and do) occur in the Standard model or its extensions: in the simplest case, the charge is transferred to leptons (which thus implies lepton number violation), but more elaborate processes, like neutron-antineutron oscillations are also possible in principle, as they don't violate electric charge conservation.

Lepton number conservation is similarly associated to phase transformations of the Lagrangian, and we know that, like in the Baryonic case, flavour violations exits via the mass terms. The question of overall lepton number conservation is however, as already mentioned, open.

Why, if lepton and baryon number play such similar roles, is the accent placed on the baryon number of the Universe rather than on its lepton number (or on matter vs antimatter)?

The answer is quite obvious, since it is in practice impossible to observe, or even less measure, the amount of neutrinos present in the cosmological background (this could however become possible some day, either via a constraint on their contribution to the mass of the Universe, or by the study of the still hypothetical Z bursts, which could result from collisions between highly energetic astrophysical neutrinos with the cosmic background, and are highly sensitive to the mass and density). For this reason, only the baryonic number of the Universe can be estimated today.

2 Baryonic number of the Universe: Why is it a problem?

From a purely empirical point of view, the very smallness of the baryon number of the Universe is problematic. Basically, a simple counting indicates the ratio of baryons to photons to lie in the window: $3\ 10^{-11} < n_B/n_\gamma < 6\ 10^{-8}$.

This number is extremely small, and prompts the double question: why is it not zero, and how is such a small number introduced (except by hand) in a theory? Further constraints, based on nucleosynthesis (which occurs late in the history of the Universe and is therefore not too sensitive to the various scenarios – even if it can be affected by the number of neutrino species and the neutrino background) indicate a stricter, but compatible bound: $4\ 10^{-10} < n_B/n_\gamma < 7\ 10^{-10}$.

These numbers, as already indicated, deal with relatively recent cosmological history. What should be the initial number of the baryon asymmetry in a "hot" Universe (by hot we mean here, at a temperature such that baryons were in thermal equilibrium). Using the hypothesis of isentropic evolution, and neglecting the masses at sufficiently high temperature so that all particles then contribute according to their number of degrees of freedom to the entropy, one gets: $\frac{n_B - n_{\bar{B}}}{n_B + n_{\bar{B}}} \sim 10^{-8}$.

Another way to view things consists in assuming that the primordial Universe developed through interactions of gravity and other fundamental forces, e.g. through the amplification of vacuum fluctuations. In such a case, gravity being blind to the difference between matter and antimatter, equal initial numbers of baryons and antibaryons are expected, and the current unbalance must be induced by subsequent interactions.

Apart from some particular mechanisms, where there is some form of explicit breaking by boundary conditions or history, such evolution thus assumes differences in matter and antimatter interactions, but also the non-conservation of baryon number at the Lagrangian level.

We should also remark that the most obvious objection, namely that the Universe could just have been created with the currently observed unbalance between matter and antimatter is itself difficult to hold in the present state of knowledge, at least in its simplest form. Indeed, we will see that quantum anomalies lead to violations of CP and of baryon number. In these conditions, an initial baryonic asymmetry would have been erased during any equilibrium period when such mechanisms were active – this would be the case of a pure baryonic number before the electroweak transition in the Standard model as it is known today. Protection of baryon number at this moment is difficult, and is one of the reasons why leptogenesis has become a favorite approach. In this latter case, lepton number is generated (way) before the electroweak transition (from a purely logical point of view, it might even be present since the onset of the universe) and converted to baryon number during the transition, assumed to take place at equilibrium.

We only mention for completeness the possibility that the observed baryon excess is a local artefact, and that the Universe is constituted with domains with either baryon or antibaryon excess. The gamma rays arising from annihilation at the boundary of such domains would be a tell-tale sign, and the fact that they have not been observed rejects such a possibility to the limit of the observable Universe.

3 Particles, antiparticles, Parity and Charge conjugation: reminders

Weak interactions break maximally the symmetry between matter and antimatter, but also break spatial parity. It turns out, as we remind in this section, that the *pure* gauge interactions respect the product of those 2 symmetries, usually referred to as CP symmetry. We discuss briefly these points in the present section, and announce already that a breaking of CP symmetry will be needed for successfull baryogenesis.

Special relativity, through the equation $E^2 = mc^4 + p^2c^2$, once transposed to the Klein Gordon equation for scalars or to the Dirac/Weyl equations for fermions, allows for any given 3-momentum both positive and negative energy solutions.

If, in many low energy problems, negative energy solutions can usually be ignored (as long as the threshold for pair creation is not met), they must be re-interpreted when addressing higher energy problems and, quantum field theory. The solution goes through the so-called "second quantisation", which re-intreprets fields not as wave functions for quantum states, but as creation and destruction operators, thereby allowing for problems with a varying number of particles.

Although very trivial, we remind here the substitutions operated, (as they are frequently obscured by simultaneous changes of variables). We thus re-interpret

- one <u>destruction operator for a negative energy particle</u> as
- one <u>creation operator for a positive energy antiparticle</u>.

In this way, the energy balance, resulting through Noether's theorem from the invariance under translations, is preserved. It becomes obvious that the same must be true of ALL converved numbers. Thus, all quantum numbers associated to antiparticles must in general be the opposite of those associated to the initially negative-energy particles. For scalar bosons, this amounts to energy, 3-momentum, and all charges (electrical, colour, weak, possibly leptonic or baryonic). For vector bosons, the spin must be added to this set. (one could notice already that the helicity is however not opposite for the antiparticle of a vector boson, as it is the projection of spin onto the direction of motion, and BOTH change sign).

The situation is similar for fermions, but includes an interesting twist. For massless fermions, one can indeed [2] use the Weyl equation rather than the Dirac one, (which is equivalent to using 2 component semi-spinors representations of the Lorentz group). Two inequivalent representations exist, one describes positive energy particles of left-handed helicity together with right-handed negative energy particles. We will refer to it as the L (for left-handed) representation. The R representation differs by the permutation of left and right-polarization.

Thus, for the L spinor, we have

- positive energy particles with left-handed (or negative) polarization
- negative energy particles with right-handed (positive polarization

When we change the language to antiparticles, both the spin and the momentum flip sign, but, as already mentioned, the helicity is unchanged:

$$h = \frac{\mathbf{p} \bullet \mathbf{s}}{\|\mathbf{p} \bullet \mathbf{s}\|}$$

The simplest representation for a fermion thus involves (assuming we take the L case).

- -one particle of negative helicity (left-handed) (lévogyre)
- -its associated antiparticle, with positive (right-handed) helicity.

Neglecting temporarily neutrino mass issues, this would describe a left-handed neutrino and its right-handed antineutrino.

It is usefull to note that this is quite particular to our 3+1 dimensional Universe. For instance, in 4+1 dimensions, this separation into L and R spinors is no longer allowed, the minimal spinorial representation has 4 components, and it is only through specific compactification schemes that the 2 component spinors are retrieved when reducing from 4+1 to 3+1 dimensions (massless chiral fermions linked to a domain wall or soliton, for instance).

Returning to our 3+1 dim world, we observe that

Charge conjugation, which consists in replacing a particle by its antiparticle, while reversing charges but not spin and momenta, is generally NOT a symmetry of the Lagrangian -or of the world *: indeed it would transform a left-handed fermion into a left-handed antifermion, which is NOT described by the same semi-spinor, and thus not necessarily present, and in any case does not need to have the same interactions.*

The situation we describe is not academi Indeed, the simplest "building bloc" for gauge interactions is composed of one vector boson and one semi-spinor, and corresponds to the very structure of the Standard Model of electroweak interactions $SU(3) \times SU(2)_L \times U(1)$ where the L subscript indeed reminds that the $SU(2)$ bosons (as was established through painstaking observation) only couple to semi-spinors of the L type - while the $U(1)$ part has specific couplings to each fermion field.

The familiar impression that parity is respected in our world, and only broken by some specificities of living organisms, is wrong, and due to the fact that, at large distance only electromagnetic forces (or at a shorter scale, atomic forces resulting from the left-over of the $SU(3)$ interactions) subsist, and that the two are indeed P conserving.

Is the lack of Charge conjugation symmetry sufficient to allow for the generation of the baryon number of the Universe? The answer is negative, and we will see why in the next section.

4 A caricatural example.

To speak in more familiar terms, we will replace in this paragraph the symmetry C (charge conjugation) by an hypothetical symmetry S, which exchanges men and women. We also use the already mentioned spatial parity symmetry (P), which here transforms left-handed into right-handed humans, and vice-versa.

To say that the world is symmetrical under S would imply only that:

- number of L women = number of L men
- number of R women = number of R men

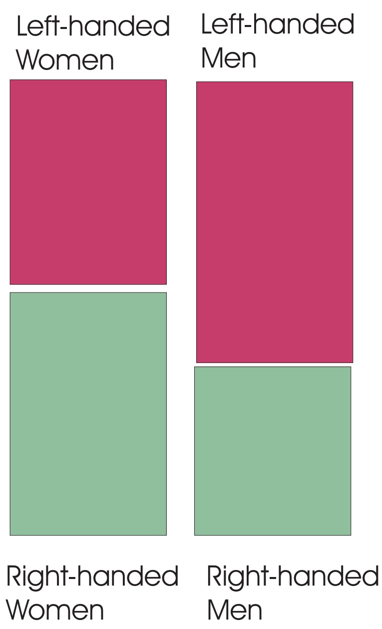

Figure 1: While S an P are not respected, SP stays a good symmetry and ensures that the total number of Men and Women are equal

while P simply states:

- number of L women = number of R women
- number of L men = number of R men

Of course S symmetry ensures also an equal total number of men and women, but its breaking is not sufficient to imply a an inequality between those total numbers. It is indeed possible to have (using obvious notations $W_L + W_R = M_R + M_L$ even if $W_L \neq M_L$ and $W_R \neq M_R$ (see Fig.1).

This happens in particular if the SP symmetry (product of S and P defined above) stays valid. It implies indeed

- number of L women= number of R men
- number of R women= number of L men

and adding the two relations yields perfect equality

$$W_L + W_R = M_R + M_L$$

The preservation of the symmetry SP (CP) was thus sufficient to preserve the equality of the total number of men and women (particles and antiparticles), even though neither S or P symmetries(or C and P) do hold.

It is in fact clear that the same is true for any operation X, such that S X (or C X) is respected. (think of replacing L and R by french and russian-speaking, for instance).

We have not completed our preliminaries yet, because an even more general symmetry, TCP, plays an important role in the discussion of the baryon asymmetry.

5 TCP and its constraints

We have just seen that CP violation was needed to generate the baryon number from an originally symmetrical Universe. We have also alluded to the fact that ***pure gauge interactions (in the absence of fermion masses or scalar couplings) are intrinsically CP-conserving*** (for details on this, see [2]). On the other hand, scalar couplings (such as fermion masses or Yukawa couplings) induce transitions between L and R spinors, and possibly CP violation. To put things in a nutshell (once again, more details are available in [2]), CP is intrinsically associated with complex conjugation at the Lagrangian level. Gauge couplings are real (which results from the unitarity of the internal groups), while scalar couplings can be complex. Therefore, the scalar couplings pertaining to a given process or its CP conjugate can differ by their phase.

We will also see later that a different process, namely quantum anomalies can induce CP violation (but only for massive fermions, so this process does not detract from the above comment which presents CP as an important symmetry of pure gauge interactions).

Nevertheless, even the introduction of complex Yukawa couplings preserves another symmetry of the Lagrangian, namely the conjugated operations CP and T (time reversal). This is known as the TCP theorem, and is valid quite generally for local interactions.

Why should we worry about TCP? In principle, this symmetry should not concern us, since there is in all cosmological problems an obvious *ex-*

plicit violation of T (and thus of TCP), due to the choice of an expanding background for the Universe.

Nevertheless, at the level of microscopic interactions, for processes much faster than the expansion, TCP remains an important constraint.

At the level of matrix elements, TCP implies permuting initial and final states,particle and antiparticle and spatial components (the latter are not mentioned explicitly in the expression below), and reads:

$$< x \mid S \mid y >=< \bar{y} \mid S \mid \bar{x} >$$

where S is the evolution operator, and $\mid y >, \mid x >$ describe the asymptotic states x et y .

As an instructive example, consider the case where $\mid x >$ simply stands for an isolated particle, x .

This allows us immediately to establish the equality between the survival probabilities (lifetimes) of the particle x and its antiparticle $\bar{x}$.

$$< x \mid S \mid x >=< \bar{x} \mid S \mid \bar{x} >$$

Comparing to the usual formulation :

$$< x \mid S \mid x >= e^{i(m+i\Gamma/2)(t-t_0)}$$

establishes that particles and antiparticles have both equal masses and equal lifetime.

There is thus no hope that the known interactions allow for instance a quicker decay of antiparticles to explain the current excess of baryons!

As a hint of an escape from this constraint, we should already remark that the constraint only applies to the total survival probability of a particle and its related antiparticle. (that is, the sum of all the possible decay channels). It does not say anything about the individual decay modes.

More explicitly, let consider a particle x with only the 2 decay processes $x \rightarrow a, x \rightarrow b$, and the charge conjugate processes, $\bar{x} \rightarrow \bar{a}, \bar{x} \rightarrow \bar{b}$. From TCP we can only infer is, for instance:

$$< a \mid S \mid x >=< \bar{x} \mid S \mid \bar{a} >$$

which relates the *desintegration* probability of x to a to the *synthesis* probability of $\bar{x}$ from $\bar{a}$.

Let us adopt the notation:

$$A_{x\to f} =< f \mid S \mid x >$$

for the amplitude, while we use P for the transition probability: $P_{x\to f}$.

Summing over all possible decay channels f, TCP implies as already mentioned, the equality of the ***total*** decay probabilities:

$$\sum_f P_{x\to f} = \sum_f P_{\bar{x}\to\bar{f}}$$

but ***does not*** imply

$$P_{x\to a} \neq P_{\bar{x}\to\bar{a}}$$

as long as this difference is compensated by other decay channels!

An almost realistic example can be given using the initially proposed baryogenesis scheme, which relied on the unification group $SU(5)$. There, heavy gauge bosons X and Y, called "leptoquarks" mediate interactions between the (unified) leptons and quarks, and can for instance have the decays (we omit Lorentz, spin and color indices):

$$\Gamma_{X\to uu} = r_u;\ n_B = 2/3;\ n_L = 0$$
$$\Gamma_{\bar{X}\to\bar{u}\bar{u}} = \bar{r}_u; n_B = -2/3;\ n_L = 0$$
$$\Gamma_{X\to e^+\bar{d}} = r_{\bar{d}}; n_B = -1/3;\ n_L = -1$$
$$\Gamma_{\bar{X}\to e^-d} = \bar{r_{\bar{d}}}; n_B = 1/3;\ n_L = 1$$

Remark in passing that these decays imply a violation of Baryon number B, lepton number L , but not of (B-L), as for instance X can decay in two channels with different baryon number. The conservation of (B-L) is just a particularity of SU(5) (and of the anomaly structure in $SU(3) \times SU(2) \times U(1)$), and in no way a general requirement like TCP.

If we compute the baryon number resulting from the decay of an initially purely symmetrical pair $X, \bar{X}$, we get:

$$n_B = 2/3\ (r_u - \bar{r_u}) - 1/3\ (r_{\bar{d}} - \bar{r_{\bar{d}}})$$

Using the equality of the $X, \bar{X}$ lifetimes, and assuming for simplicity now that these are the only decay channels involved, we also have, by TCP

$$r_u + r_{\bar{d}} = \bar{r_u} + \bar{r_{\bar{d}}}$$

which leads to :

$$n_B = r_u - \bar{r_u}$$

We "only" need to ensure that $r_u \neq \bar{r_u}$ to generate a non-vanishing baryon number from an initially symmetrical Universe, and this, despite the local use of TCP.

How can such a disparity between the two decay rates be obtained? We send again for more details to the reference ([2]), and sketch the basis of the mechanism in the next section.

6 Channels compensation : Reconciling baryon asymmetries and TCP

As should appear clearly from the previous section, we need not only C and CP violation, but also a difference between the partial decay rates of C or CP conjugated particles. It should also be clear from the above evocation of TCP that such difference can only exist if decays are permitted through more than one channel, and if, in some way, each of these channels is "aware" of the others, so that compensations can occurs, ensuring that the lifetime of a particle and its charge conjugate stay the same.

From the figure 2, it is quite obvious that this cannot happen at first order: each channel appears as a separate amplitude, and ignores the others (it is easy to check that CP conjugate particles have the same partial branchings at first order. What we illustrate further is the case where 2 channels interfere - let us call them $X \rightarrow a$ and $X \rightarrow b$.

At second order, the final state a can be reached either directly, or through an intermediate step, $X \rightarrow b$, and a later rescattering $b \rightarrow a$. The two processes will of course interfere, and this brings the necessary exchange of information: channel a is now aware of the existence of channel b , and compensation between the partial decays can occur, so that $\Gamma(X) = \Gamma(\bar{X})$ while keeping $\Gamma(X \rightarrow a) \neq \Gamma(\bar{X} \rightarrow \bar{a})$.

Let us make this slightly more explicit. In the simple case of a scalar X decaying through complex Yukawa couplings λ_a, λ_b into channels a, b, the couplings of $\bar{X}$ are simply complex conjugates. At first order, only $|\lambda_a|^2$ intervenes for the decays into channel a (or $\bar{a}$), and no difference can arise. At the next order (third order in λ) we must include a rescattering term between the 2 channels. We write, for the rescattering $R_{b\rightarrow a}e^{i\alpha}$ where R is real, and α is the phase associated to the Yukawa couplings appearing in the vertices. Quite obviously, the charge conjugate process has opposite phase: $R_{b\to a}e^{-i\alpha}$. This is however still not sufficient (as is easily checked).

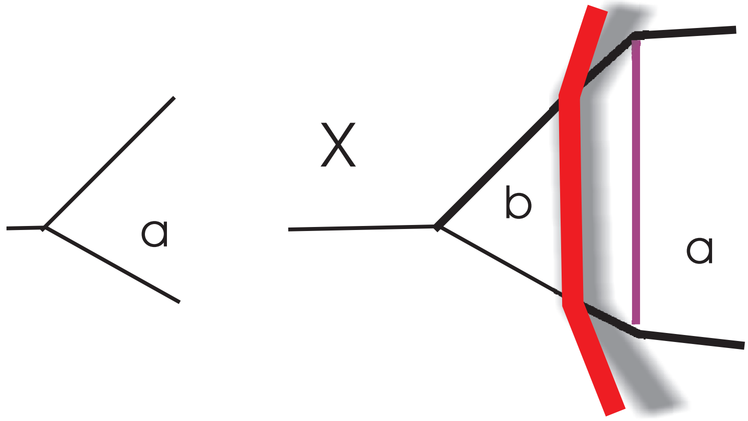

Figure 2: interference between channels a and b

In some way, the process must know that the intermediary state (here, the channels b or $\bar{b}$) are actually open (that is physically realizable), and not simply virtual states, for a compensation to be possible. This is indeed the case, and the presence of an intermediary *physical* or *on-shell* state is well-known to introduce an imaginary part in the Feynman amplitude. This is usually exhibited by writing all the possible "unitarity cuts", where all the "cut" lines must be simultaneously on-shell. We represent the presence of this imaginaly part by $e^{i\xi}$. It must be noted that this phase is present only for unitarity reasons, and only depends on the mass (in particular, the phase space), and not on the nature of the particles or antiparticles. Thus, the phase ξ is insensitive to the fact that we start from X or $\bar{X}$, and does NOT flip between the 2 processes. We thus get:

$$\Gamma(X \to a) \sim |\lambda_a + \lambda_b e^{i\alpha} R_{b\to a} e^{i\xi}|$$
$$\Gamma(\bar{X} \to \bar{a}) \sim |\lambda_a + \lambda_b e^{-i\alpha} R_{\bar{b}\to\bar{a}} e^{i\xi}|$$
$$\Gamma(X \to a) - \Gamma(\bar{X} \to \bar{a}) \sim \lambda_a \lambda_b R_{b\to a} sin(\alpha) sin(\xi)$$

The latter relation clearly shows the intricate conditions required to get different decay modes for particles and antiparticles, despite the CPT theorem: need for compensating channels, need for them to be kinematically accessible, need for CP violation (the phase ξ).

We must furthermore remark, in preparation for the next paragraph, that we have this far assumed a decay "in vacuum". This is quite unlikely, and we must expect that, at least in the early Universe, the decay will occur in some form of thermal bath. We must thus ensure that the reverse reactions does not negate the desired effect of asymmetry between particle and antiparticle fate.

For this, the condition is that the decay process (or other processes generating the baryon number) occurs out of equilibrium.

Note that all the points relative to generation of baryon number above also apply to lepton number – as we shall see below, the leptogenesis mechanism precisely relies on initial generation of lepton number, later followed by its conversion to baryon number.

To summarize things in a nutshell, we have shown in this section that particles and antiparticles can die in different ways, despite having the same lifetime!

7 Sakharov's conditions

We have under way met with the 3 conditions for baryo (or lepto-) genesis, better known as Sakharov's conditions:

- violation of baryon (- lepton) number
- violation of C and CP symmetries
- the process must occur out of equilibrium

Since the pioneering work of Sakharov [3] and Yoshimura [4], numerous models have been suggested. We will not review them in details, but will consider in the following sections various mechanisms used to satisfy the individual conditions above.

We will then put those mechanisms together to describe more specifically one of the favored schemes, namely baryogenesis through leptogenesis.

Note that some other scenarios are possible, which in some way evade the conditions above (for instance, a baryon-number scalar develops vacuum expectation value during the cosmological evolution of the Universe) [5]; we will however not consider them here.

8 Baryon and / or lepton number violation mechanisms

If generation of the Baryon number of the Universe were the only rationale for introducing baryon number violation in the model, the intellectual gain would be far from obvious. Fortunately, this is not so, as Baryon number violation occurs automatically in theories of grand unification (by the very fact that quarks and leptons need to be introduced in the same representations). In such cases, baryon and lepton number are usually linked. Other specific mechanisms exist for Lepton number violation (see later).

Quite interestingly, baryon and lepton number violations also appear in the Standard model, quite independently of the unification (see below: anomalies).

For the moment, we will concentrate on the baryon and lepton violations linked to grand unification.

The Standard model, based on the gauge group $SU(3) \times SU(2)_L \times U(1)$ does not really unify fundamental interactions, even if it provides them with a common gauge structure: indeed several coupling constants are still present, in particular for the abelian part of the group. While anomalies can put some restrictions on these couplings, it is quite likely that their cancelation in fact stems from unification in a single (semi-simple) group.

Trial and error has shown that the smallest practical such group is SU(5), with the fermions placed in 5 and $\bar{1}0$ representations (for each family, and assuming no "right-handed' neutrino is present - the latter would need including a singlet).

A more elegant unification, including all fermions of one family (including the still hypothetical ν_R) in a single representation relies on using the 16 of $SO(10)$.

In all such cases (or in even more ambitious unification schemes, but with the above cases as subgroups), baryon an lepton number violation will take place (for instance through the process $u+u \to X \to \bar{d}e^+$ already mentioned.

What remains to be explained is the extraordinary protection needed for the proton lifetime.

While some specific mechanisms may be at play (for instance specific quantum numbers introduced by hand in supersymmetry), the basic tool is to impose a very high mass for the intermediary boson responsible for this breaking. (X in the above example). It must be noted that this high mass constraint is obtained independently of the arguments based on the running of coupling constants, which also suggest a very high unification scale.

What are the orders of magnitude? For a particle of mass m with allowed decay and no suppression, one has,

$$\Gamma \simeq \kappa m$$

, which, for $m = 1\text{GeV}$ and $\kappa = 1$ leads to $\tau \simeq 6\ 10^{-25}s$

If, instead of a "strong" style of interaction, one uses an intermediary vector boson of mass M_X, a factor ${M_X}^4$ appears in denominator, and must be compensated dimensionally, which, together with a coupling constant factor g^4 leads to

$$\Gamma \simeq g^4 m^5 / {M_X}^4$$

It is then a matter of choosing M_X large enough to move from the initial lifetime ($10^{-24}s$) to the observed limit $\tau_{p\to\pi e^+} > 10^{32}$ years !

This leads us to the usually accepted grand unification range ($10^{16}GeV$)

We will not go into further detail for the time being, except to mention an "accidental" characteristic of $SU(5)$ grand unification. As can be checked in the particular example given above, while Baryon number B and Lepton number L are separately not conserved, the difference $B - L$ is conserved. This is however a peculiarity of $SU(5)$, and this symmetry is instead part of the gauge symmetry of $SO(10)$; it is broken in the transition $SO(10) \to SU(5)$.

9 Quantum Anomalies: When the quantum world ignores classical symmetries.

Continuous Lagrangian symmetries imply current conservation through Noether's theorem. Typically, if ψ represents a fermion field (or a multiplet of them), the invariance of

$$L = \bar{\psi_L} D^\mu \gamma_\mu \psi_L$$

under $\psi_L \rightarrow e^{i\alpha}\psi_L$ implies the classical conservation of the current

$$\partial_\mu {j_L}^\mu = 0$$

where

$${j_L}^\mu = \bar{\psi_L} \gamma^\mu \psi_L$$

The same would obviously be written for a possible R component.

It came however as a surprise that this conservation does not hold in the quantum world, where so-called "quantum anomalies", like the presence of a triangular diagram connect the fermion sector to specific gauge field configurations, inducing a non-vanishing of the divergence. While surprising at first, these anomalous terms are unambiguously present, and their effect is tested in radiative decays of mesons. They stem from the regularization of linearly divergent integrals, but are by themselves perfectly finite and well-defined.

Such anomalies are on one hand necessary in some non-gauged currents (like the axial current associated to the pions) to explain experimentally observed decays, but on the other hand cannot be accepted in gauged currents, where they would impair the renormalisation of the theory.

As a matter of fact before grand unification, in $SU(3) \times SU(2)_L \times U(1)$ the U(1) charges must precisely be adjusted to avoid such anomalies (this is automatically realized in SO(10), which does not present anomalies, and as a consequence in SU(5) with the usual representations, which appears as a subgroup of the former).

As already mentioned, such anomalies may subsist for those currents which are not "gauged". This is precisely the case of the Lepton and Baryon number: no long-range interaction is associated to those numbers, despite the fact that they are remarkably well conserved in our obervable surroundings. If we neglect mass terms, this is in particular even true for the total number of L baryons (to which are associated R antibaryons) or for the lepton numbers (see however a dedicated section below).

We can for instance write

$$\partial_\mu j^\mu_{lepton,L} + \partial_\mu j^\mu_{baryon,L} = \kappa \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}$$

where the right side of the equation refers to SU(2) gauge fields.

This equation shows how a baryon charge can be exchanges for a lepton charge and a change in field configurations.

Such a mechanism for violating conservation of quantum numbers is rather general, and was evoked already by 't Hooft in the context of strong interactions and instantons. The efficiency of such mechanisms is however in general very low, as it requires to excite topologically non-trivial configurations, (which appear in a non-perturbative way and are non-local) from the usual local particle fields.

At low temperature, 't Hooft estimated that such effects should be suppressed (for the weak group by a factor $e^{-4\pi/\alpha_W}$).

In the context of spontaneous breaking of the electroweak symmetry (Brout-Englert-Higgs mechanism) operating in a cosmological context (that is in a thermal bath), it has been argued by Klinkhammer and Manton [6] that unstable solutions, (named sphalerons), corresponding to a potential barriers between vacua of different baryon and lepton numbers would appear, with a mass comparable to the temperature of the transition. The transition probability is then considerably increased, to $e^{-M_{sphaleron}/kT}$, which means that it approaches unity for $T \sim M_{sphaleron}$ (the latter mass is of the order of the electroweak transition energy, namely 100 GeV).

We are thus, if this scheme is correct, presented with an almost ineluctable mechanism for generation of baryon or lepton number – but also for their destruction, should the transition occur at or close to equilibrium.

We should be somewhat more specific. First of all, to precise that the quantity directly affected by this spaleron mechanism is actually $(B+L)_L$, (since only the Left–handed particles are connected to the SU(2) group), and Right-handed components are only touched through their indirect coupling through the mass terms. As in SU(5), the (B-L) current is conserved by the process. This will reveal to be of importance (and in fact, catastrophic for the schemes based on the decay of SU(5) heavy intermediaries).

A word of caution is however in order. The existence of sphaleron solutions has only been rigorously demonstrated for the group $SU(2)$, and numerical evaluations have extended it to $SU(2) \times U(1)$, always in the broken phase, where the vacuum expectation of the scalar field provides the dimension of the sphaleron energy. It is however frequently advocated that similar configurations are active around the phase transition.

The efficiency of the mechanism is also unclear. Evaluations must take into account the extended character of these configurations, which are not necessarily easy to excite from particle states. In particular, estimations of the efficiency of this mechanism out of thermal equilibrium don't have readily measurable equivalent, except for numerical simulations.

More reliable probably is the assumption that, if the phase transition occurs slowly (second order phase transition), not far from equilibrium, the process will have time to complete to saturation. The risk then is quite high to see any $SU(5)$-generated baryon number destroyed. In this context indeed, any SU(5) decay will respect the $B-L$ symmetry, so before the weak transition $B-L=0$ even while $B, L \neq 0$. The phase transition at equilibrium will destroy any $B+L$ while keeping $B-L$ constant: thus no baryon number can in principle survive.

The only ways out are

- Assume that the electroweak phase transition generates itself a non-vanishing B - it becomes then irrelevant whether previous baryogenesis did occur. This however requires an out-of -equilibrium transition, which seems for the moment excluded for the known parameter of the Standard model, and extensions are needed (extra scalar fields could do the job, and are present for instance in supersymmetry; even singlets fields could suffice, by providing trilinear couplings. Notice however that this electroweak baryogenesis remains extremely sensitive to the details of the phase transition, to the particularities of the sphaleron mechanism, and requires important CP violation at low energy, which is not normally found in the Standard Model (see below)

- Assume instead that the electroweak phase transition occurs close to equilibrium, and merely redistributes the values of B, L, assuming that before the transition $B-L \neq 0$. Since this number is conserved through the transition, nonzero B will in general emerge. A particular (and popular) case is "leptogenesis", where the high temperature mechanisms are assumed to generate only L, later to be turned in to B. The popularity of this mechanism is largely linked to its insensitivity to the details of the electroweak phase transition and sphalerons, provided the system stays to equilibrium long enough for the transformation to approach saturation.

10 Specific sources violating Lepton number

While baryon number violation in $SU(5)$ is merely a transfer from baryons to leptons present in the same multiplet, through gauge boson exchanges (and conservation of $B - L$), the leptonic sector, either in grand unification schemes or simply in the $SU(2) \times U(1)$ framework, offers room for a more direct violation of lepton number.

The simplest case is to consider the right-handed neutrinos, which are singlets in $SU(2) \times U(1)$ and $SU(5)$ (but are part of the 16 representation in $SO(10)$). Let us mention in passing that lepton number violation is also possible without introducing ν_R, as a Majorana mass term for the ν_L may be introduced in the Standard model, but at the cost of including complex scalar triplets, with a vacuum expectation value small enough not to upset the $m_W/(M_Z cos(\theta))$ value.

Gauge interactions impose current conservation, and in general invariance under transformations of the type $\psi_{L,R} \rightarrow e^{i\alpha}\psi_{L,R}$, and, for the conjugate field $\overline{\psi_{L,R}} \rightarrow e^{-i\alpha}\overline{\psi_{L,R}}$, which are compatible with the "Dirac" mass terms

$$L_{cin} = \bar{\psi_L} D^{\mu} \gamma_{\mu} \psi_L$$

$$L_{Dirac} = m \bar{\psi_R} \psi_L + h.c. \ ,$$

But similar constraints don't apply usually to the right-handed neutrino ν_R, which is a singlet. (In SO(10), the right-handed neutrino is not a singlet, and the Majorana mass appears through a breaking of the gauge symmetry). In fact, in $SU(5)$ or in the standard model, the right-handed neutrino is essentially decoupled from the other particles, its only interaction being confined to the mass term which links it to active neutrinos.

As the only requirement for a mass term in the Lagrangian is to be invariant under Lorentz transformations, one can thus introduce a "Majorana mass term", which, in terms of 2-components spinors reads:

$$L_{Maj} = M \epsilon_{ij} {\eta_R}^i {\eta_R}^j + h.c.$$

(remember that the fermion spinors are anticommuting fields). It is quite obvious from this expression that the coupling does not respect any phase transformation, and in particular that leading to the conservation of ν_R number.

It is often more convenient to write the above coupling in Dirac notation (although this is redundant for 2-component spinors, it is usefull when we have to deal both with 2 and 4-components fields): introducing the notation ${\psi_R}^c = C\psi^{+t}$, with C the charge conjugation matrix

$$L_M = m\overline{{\psi_R}^c}\psi_R + h.c.$$

Not only is ν_R number violated, it also becomes impossible to speak of particle or antiparticle: the ν_R becomes thus its own antiparticle.

As long as the ν_R is not coupled to the usual particles, such a term of course has no consequence, but the presence of a mass term (or, in practice, a Yukawa coupling) between ν_L and ν_R transfers this violation to the usual leptonic fields:

$$L_{Yukawa} = \frac{m}{v}\overline{\psi_R}\widetilde{\phi}^\dagger\Psi_L + h.c.$$

where Ψ_L stands for the electroweak doublet (ν;e)

When $\widetilde{\phi}$ develops a vacuum expectation value ($v/\sqrt{2}$; 0) this results in a mass term, and one faces a mass matrix of the type:

$$\begin{pmatrix} 0 & m \\ m & M \end{pmatrix}$$

The diagonalisation of this matrix leads to mass eigenstates with approximate masses (assuming $m \ll M$) m^2/M and M - a mechanism known as "see-saw".

The mass term generated for the "light" neutrino is also a source of lepton number violation: it should in principle be observable. In practice, most manifestations of neutrino mass terms are "inertial mass", as they enter through phase space, or the energy-momentum relation: in this case, the Majorana mass term is indistinguishable from a more usual Dirac mass. The only process where we can hope to observe the effect of lepton number non-conservation at low energy is in practice the neutrinoless double beta decay (of course processes involving μ or τ could in principle be considered, but they are forbidden by lack of phase space in nuclear decays, and it would be impossible to reach the required sensitivity in other experiments).

$$N \rightarrow N' + 2e^-$$

This process is of course the object of very active experimental investigation.

We have thus introduced a specific violation of lepton number L, which does not affect B. As a result, not only L but also $L, B - L, B + L$ are affected.

This violation is stealthy in the current state of the Universe, but becomes manifest at high temperature, when T is comparable to the mass M of the heavy (mostly right-handed) neutrino, which we will now call N.

N being its own antiparticle, can decay both into letpon and antilepton channels, namely (we spell out very explicitly the nature of the leptons – indeed we must remember that the antiparticle of a Left-handed electron is a r Right-handed positron).

$$N \to e_R + anti(e_L) + \nu_L$$

$$N \to \bar{N} \to anti(e_R) + e_L + anti(\nu_L)$$

We have thus put together some of the elements (previously introduced in the framework of baryon number violation), namely the existence of L violation, and the possibility of competing channels for the decay of the N particle, necessary to overcome the constraints of TCP (see above).

11 Losing balance (equilibrium)

As mentioned previously, the benefits from C, CP and B or L violation required for generating a non-vanishing L or B are lost if:

- the transitions (e.g. decays) supposed to generate Baryon or Lepton number occur at or close to equilibrium;
- or if, B being created at high energies with $B - L = 0$, the electroweak phase transition occurs later at or close to equilibrium, creating the conditions for washing out the previously obtained excess.

We now list some possible situations where the desired departure from equilibrium could be found.

11.1 Relic particles

We first consider the mechanisms proposed by Sakharov and Yoshimura,[3] [4] namely, the B (or L) and CP- violating decay of a heavy particle.

Since the particles are very massive (much more than the weak scale, at least) we must return to the cosmological period where the temperature was high enough that such particles could be abundant, in equilibrium with a thermal bath of temperature $T \geq M$. We can then reasonably assume that their equilibrium density is reached, and is given by $e^{-E/kT}$.[1] When the Universe cools down, this density SHOULD decrease, but this requires some mechanism (annihilation or decay). It may happen that such mechanisms are too slow to keep pace with the cooling (and expansion) of the Universe. In this case, the population of particles stays much higher than the naïve thermodynamical expectation, and we speak of "***relic particles***".

Such particles are particularly interesting for our purpose. After surviving the cooling of the Universe, their decay at Universe temperatures much lower than their masses, produces secondary particles (typically the known leptons or quarks) with energies much higher than the ambiant thermal bath. As a result, the inverse process (recombination of the products to re-build the initial heavy particle) becomes highly unlikely, and the decay is completely "out of equilibrium".

To get an idea of the orders of magnitude involved, we consider the simple case of the desintegration of a relic particle of mass M, assuming simply a 2-body phase space and a coupling g.

The decay rate is then typically given by $\tau^{-1} = \Gamma \cong g^2 M$ and should be compared to the expansion rate of the Universe at the time (or temperature T) of decay, given by the Hubble constant H. We need:

$$\tau \gg H^{-1}$$

The value of H is given at high temperature by $H = \sqrt{g^*}T^2/10^{19}GeV$ where g^* counts the effective degrees of freedom available at temperature T.

Taking $T = M$ to characterize the decay at the time the particle falls out of thermal equilibrium, we get:

$$M \geq \frac{g^2}{\sqrt{g^*}} 10^{19}GeV \sim 10^{16}GeV$$

It is a striking coïncidence that the scale obtained by this "out-of-equilibrium" criterion (assuming the particle is "typical", i.e., that its decay is not extraordinarily suppressed, as can be the case in very specific models), is very similar to the "grand unification scale" already mentioned as a possible source of B

[1] Just a side note here: in the case of inflation, the initial density of particles is diluted by the expansion, and becomes negligeable. The following re-heating mechanism may later couple more strongly to some particles than others, possibly resulting in non-thermal distributions, particularly for very weakly coupled sectors

or L violation. Note that this grand unification scale is itself determined by two independent considerations (actually in slight conflict in the case of the minimal $SU(5)$), namely, the convergence of the running coupling constants, and the lower limit inferred from proton stability requirements.

This out-of-equilibrium decay mechanism has been largely used, in a variety of schemes. The most obvious (and currently favoured one) is the decay of a very heavy "lepton" field (typically a Majorana right-handed neutrino usually noted N). Depending on the mass, even a sizeable coupling can make this a relic particle (for instance in $SO(10)$, where $SU(2)_R$ bosons are present, but very weak couplings (if the R breaking scale is much higher than m_N) can also occur if only the Yukawa terms linking this particle to the light fermions contribute to the decay. In this case, the N mass scale can be brought down.

There are a number of cases, as a matter of fact, where the relic character of a particle arises not primarily from its high mass, but from a suppression of its coupling to potential decay channels. Such is the case, for instance, for supersymmetric partners: the lightest supersymmetric particle is usually protected from decay by some ad-hoc R parity. A small breaking of this parity then accounts for a very slow decay rate.

More exotic mechanisms may even be drummed up: for instance, relic particle can stay trapped for a long time in singularities ("cosmic strings"), where they are effectively massless. Upon the late evaporation of these singularities, the particles are released with a mass larger than the current temperature.

11.2 Phase transitions

Another possible source of out-of-equilibrium processes comes from phase transitions. A close analogy is provided by boiling water, where a bubble of "true vacuum" (here vapour, the favoured state at high enough temperature) develops in a medium which has become unstable, and expands in an irreversible manner.

In the cosmological framework, the phase transition is supposed to happen during the cooling of the Universe, and could be associated for instance to the electroweak transition (there may be many successive phase transitions in a grand unified theory, but we will focus on the last): the false vacuum corresponds then to the unbroken phase, while the true vacuum (which we live on) sees a developing vacuum expectation value for a scalar field (Brout-Englert-Higgs field).

One proposed mechanism uses the expansion of this bubble, and the differential reflection of fermions on it: for instance, top quarks outside the bubble would be massive, but would acquire a heavy mass inside: the lowest energy ones could then not penetrate the bubble. Many variants of this (or similar) mechanism have been suggested; in addition to the above ingredients, they must of course include baryon number violation (unsuppressed at the phase transition, according to the sphaleron approach), and CP violation (usually not suffiecient in the Standard model at such energy).

More importantly for our current discussion of equilibrium, for the above mechanism to work, the transition needs to be of "first order", and followed by a fast cooling, to make sure that the process does not come into equilibrium. It has been shown that, in the strict context of the Standard model (only one doublet of scalars), this led to unacceptable constraints (namely, a mass of the Brout-Englert-Higgs scalar of 50 to 60 GeV, which is completely excluded by LEP data.). It should be kept in mind however that even minimal variants of the model might be reconciled with the first order phase transition, for instance if scalar singlets or triplets are introduced, leading to trilinear couplings, or in supersymmetric extensions. This approach would obviously benefit from experimental support, which may come with the LHC.

For the time being, we will return to the default assumption (the minimal scalar structure for the Standard model), which leads, with the current constraints on the scalar mass, to a much smoother second order phase transition. In such a case, the baryon number violation associated to anomalies (and sphaleron-type solutions) operates close to equilibrium, and tends to be complete, that is, to obliterate completely $(B + L)_L$ (remember that in the Standard Model, the sphalerons act on the left-handed fields, and conserve $B - L$). This could have the unwanted effect of wiping out any previously generated baryon number with $B - L = 0$ (an example of which is the baryon number generated by decay of heavy particles in $SU(5)$).

A contrario, this mechanism may be used in other schemes, notably leptogenesis, to tranfer the initally generated L asymmetry to the baryonic sector, doing so in nearly complete way (and thus without need to compute the details of the difficult to describe phase transition).

12 How can we break CP?

This is probably the hardest question to answer as of today. As mentioned before, CP is the natural symmetry of *pure gauge theorie*, that is if no scalar interactions (including mass terms) are introduced. The source of CP violation must thus be found in scalar couplings (fundamental or effective), or complex vacuum expectation values (in the case of spontaneous CP violation).

Unfortunately, we have no rules to constraint this scalar factor, and in most cases, the CP violation responsible for lepto- or baryogenesis is introduced in a pure "ad-hoc" way. In the best case, it might be hoped that such "ad-hoc" CP violation might be related to low-energy observables, bringing at least some constraints, but this can usually only be done at the cost of further assumptions. (for instance, assuming spontaneous CP violation in Left-Right symmetrical models, or betting on some particular "texture" of the lepton masses).

The most obvious question of course (particularly before lepton mixing and the possibility of CP violation in leptons were established) is whether the currently established CP violation in hadrons could in fact be, just by itself, responsible for baryogenesis. This way has been explored by a number of authors, but is in fact rather hopeless. Relying on the Kobayashi-Maskawa mechanism, CP violation in the K and B systems calls indeed into play the 3 generations of quarks, and the 3 mixing angles of the KM matrix, on top of the CP violating phase. This is made particularly clear by the approach of the Jarlskog invariants [7], and the expected effect depends on:

$$\begin{aligned} J &= sin(\theta_1)sin(\theta_2)sin(\theta_3)sin(\delta) * P_u * P_d \\ P_u &= ({m_u}^2 - {m_c}^2) * ({m_t}^2 - {m_c}^2) * ({m_t}^2 - {m_u}^2) \\ P_d &= ({m_d}^2 - {m_s}^2) * ({m_b}^2 - {m_s}^2) * ({m_b}^2 - {m_d}^2) \end{aligned}$$

Quite obviously, this quantity has a high mass/energy dimension, (GeV^{12}), and non-dimensional CP violating effects (ratios) require proper normalisation. In the case of the K system, some small dimensional parameters are available (like the K_L - K_S mass difference) and furthermore, part of the constraints (some mass factors simply enforce the possibility to distinguish the various quarks) are fulfilled by external conditions, hence a large ratio (but not a large effect in absolute terms) can be obtained.

In the present case of baryon number generation however, things happen usually at a much higher energy, for instance $100GeV$ for the electroweak transition. The mass differences of the light quarks are then inoperative, as is seen by scaling the determinant J by the transition energy, leading to an effect less than 10^{-17}, which is totally insufficient in for the desired baryon excess.

Our only hopes for the moment to elucidate the source of CP violation in baryogenesis, are on the one hand a fundamental understanding of the origin of CP violation (for instance in the compactification mechanism associated to the dimensional reduction of a fundamental gauge theory in extra dimensions), or to hope for some low-energy signal of CP violation beyond the Kobayashi-Maskawa scheme: this could conceivably be the detection an electric dipole moment for the neutron (above the tiny value expected in the Standard model), or CP violation in the leptonic sector.

We will not speculate further, and assume in the scenarios discussed below that CP violation is introduced "as usual" , that is, by ad-hoc Yukawa couplings.

13 Some possible schemes for baryon number generation

With the "building blocs" in hand, we now turn to some possible scenarios (the currently favoured case of leptogenesis will be dealt with specifically in the next section).

The most direct approach is that of Sakharov and Yoshimura. Namely, a very heavy particle (for instance a "leptoquark" boson of SU(5), or a scalar particle of similar mass) is assumed to become a relic particle before it decays asymmetrically, as discussed above. The main drawbacks of this mechanism come from 2 different sources. First, we must assume a completely ad-hoc and in practice untestable mechanism for CP violation, acting at temperatures close to the unification scale. The second criticism is more specific to SU(5), since models based on this group conserve $B - L$. Any baryon or lepton number generated through such mechanisms will thus satisfy $B - L = 0$.

This brings trouble at the electroweak phase transition, since in the simplest case of the Standard model, this occurs close to equilibrium, so that the mechanisms associated to anomalies and in particular sphaleron solutions tend to bring $(B + L)_L \to 0$. The right-handed components are also affected through mass terms, and brings the system to $B - L = 0 = B + L$, the first equality being due to the specifics of SU(5), the second to the mechanism of electroweak transition. This is equivalent to a complete wash-out of B and L.

Knowing the illness is of course here finding the cure: we need to find a scheme where the machanism of Sakharov and Yoshimura generate a non-vanishing $(B - L)$. The currently favoured scenario in this direction is precisely Leptogenesis, which we will study in the following section.

Other schemes, usually more speculative, try to generate the baryon number at the time of the electroweak phase transition. As alluded to before, if this attempt is at first sight tempting, since in a way all the ingredients needed (out-of-equilibrium stage due to phase transition, CP violation due to the Kobahashi-Maskawa matrix, and baryon number violation due to anomalies). However, we have seen above that the transition is not out-of-equilibrium in the minimal Standard model once the current bound (113 GeV) on the mass of the scalar Brout-Englert-Higgs boson is taken into account, and that the CP violation invoked is too small by several orders of magnitude.

It is of course possible to circumvent those difficulties, at the cost of complicating the model. As already mentioned, additional scalar fields (at least an additional doublet is in any case needed in supersymmetric extensions), possibly including singlets or triplets, would allow for a first-order transition, even with the current lower bound on the scalar mass. More arbitrariness comes from the CP-violation mechanism to be invoked, (for instance, a hord of parameters appear in supersymmetric extensions).

We should also mention here completely different approaches, less related in a way to the details of fundamental particles interactions than to cosmological models. A typical example is the Affleck-Dine mechanism,[5] based on the fluctuations of a primordial scalar field, carrying lepton or baryon number. As this approach is quite different from the main theme pursued here, we simply refer the reader to ref. [1] for a more exhaustive review.

14 Leptogenesis

We devote now an important section to the currently most popular model of baryogenesis, based in fact on an initial violation of lepton number.

It is difficult to pinpoint the reason for the current popularity of this approach, which has in fact been considered for quite some time.[8] Probably the recent acceptation of neutrino oscillations as a fact, and the subsequent popularity of the "see-saw" mechanism to explain the smallness of the neutrino masses is an important effect. Intense work to extract possible low-energy consequences by Buchmüller and Plumacher [9] is also certainly also a factor, but probably the stronger point of the model is that, while incorportating the non-perturbative violation of baryon and lepton number by anomalies, and a conversion mechanism based on sphalerons, it does not depend crucially on the details of the electroweak transition - provided it takes place close to equilibrium. In that way, the approach is finally on sounder ground than many.

The basic scheme is thus simple: at high energy, heavy Neutrinos with Majorana masses become relic particles, which decay asymmetrically into light leptons and anti-leptons. The lepton number violation is present due to the Majorana character of those relic particles (see above), but the CP violation has to be put in by hand in the Yukawa coupling between scalar fields, heavy and light fermions. The out-of-equilibrium condition is fulfilled by the relic character of the particles, which only places mild constraints on their coupling, provided the mass is taken to be high enough.

This leaves us to face the electroweak phase transition with no baryon number and a net lepton number. This time, use is made of the near-equilibrium transition to effectively convert part of the lepton number of the Universe into baryon number. If B_t and L_t stand for the corresponding baryon and lepton numbers at time t, and L_0 the inital lepton number, during the phase transition, we must keep (remember that $B - L$ is conserved by the Standard model)

$$B_t - L_t = -L_0,$$

Assuming that the transition stays close enough to equilibrium, it will tend to achieve

$$B_t + L_t \rightarrow 0,$$

so that upon completion of the transition $B_t = -L_0/2$.

This qualitative description is substantiated by (much) more complicated evaluations [10] which yield $B_{final} = -28/79\ L_0$, quite close to the naïve estimation.

Having sketched the basic framework, we now turn in the following subsections to some details of the mechanism

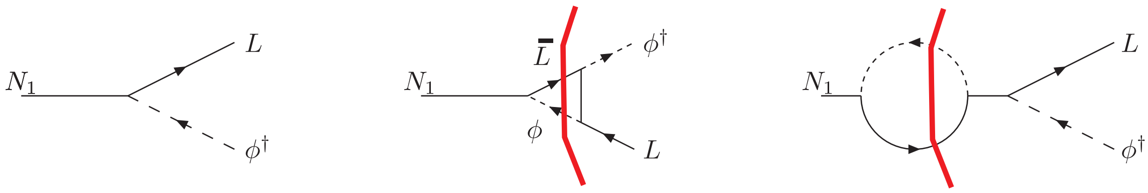

Figure 3: L-violating decay of N_1 Majorana neutrino generating the CP asymmetry in leptogenesis; the interfering channel with opposite lepton number is shown, as are possible unitarity cuts.

14.1 Generation of the Lepton number

Let us concentrate temporarily on the Yukawa coupling between a (right-handed) singlet neutrino N and a lepton doublet L

$$\lambda N \phi^\dagger L$$

where λ is the arbitrary coupling (in fact, a matrix in lepton family space), and ϕ is the usual scalar doublet.

The basic mechanism for generating lepton number in the decay of the relic particle has been reviewed before. Here we show the relevant graph in figure 3 for the decay $N \to L\phi^\dagger$, while the compensating channel $N \to \bar{L}\phi$ appears through the unitarity cut in the triangle (of course, we should show the corresponding graphs for the "compensating" channel - we keep using this expression here, although N is its own antiparticle, because the diagrams are exactly the same as in the heavy leptoquark decay, and therefore the compensations occur in the same way).

One peculiarity here is the emergence of the "bubble" diagram on the initial N line. This diagram must be included, as it can contribute and is of the same order in perturbation in the Yukawa coupling. For those who would be concerned with the (formal) inclusion of a one-particle reducible Feynman diagram in our evaluation, suffice it to say that the same contribution would appear after as a counterterm in the definition of the mass, after a proper substraction scheme. A very similar situation was met in a totally different context when computing in a gauge invariant way the decay of K mesons into axions,[11] and its importance in the present context (particularly when the N fermions pertaining to different generations can be nearly degenerate) was stressed in [12].

The light lepton asymmetry resulting from this channel is given by

$$\epsilon_i^\phi = \frac{\Gamma(N_i \to l\,\phi) - \Gamma(N_i \to \bar{l}\,\phi^\dagger)}{\Gamma(N_i \to l\,\phi) + \Gamma(N_i \to \bar{l}\,\phi^\dagger)},$$

We will assume for simplicity that the heavy N are well-separated in mass, with $M_1 << M_2 << M_3$, in which case it is easy to see that the generated lepton number is associated with the decay of the lightest state, N_1. Taken alone, this decay mechanism leads to an asymmetry in the (light) lepton number given by;

$$\epsilon_1^\phi = -\frac{3}{16\pi}\frac{1}{\left[\lambda_\nu\lambda_\nu^\dagger\right]_{11}}\sum_{j\neq 1}\mathrm{Im}\left(\left[\lambda_\nu\lambda_\nu^\dagger\right]_{1j}^2\right)\frac{M_1}{M_j}.$$

It is usefull to define the parameter

$$\tilde{m}_i = \frac{v^2\left(\lambda_\nu\lambda_\nu^\dagger\right)_{ii}}{M_i}$$

We need however to remark that this parameter is *not* directly related to the light neutrino masses, although it appears very similar and has the same dimensions. Any relation between $\tilde{m}_i$ and the observable neutrino mass m_1 depends thus highly on the details of the mass pattern assumed (texture of the M and λ matrices).

Even the simple case considered above contains far too many parameters for our purpose, and considerable effort has been given to establishing at least upper bounds for ϵ_1. Davidson and Ibarra first deduce the following upper bound [13]:

$$|\epsilon_1^\phi| \leq \epsilon_{DI}^\phi = \frac{3}{16\pi}\frac{M_1}{v^2}\left(m_3 - m_1\right).$$

As an estimate, taking e.g. $M_1 = 10^8 GeV$ and $m_3 = \sqrt{\Delta m^2_{atm}}$, $m_1 = 0$, the bound yields $\epsilon \sim 10^{-8}$, allowing a baryon asymmetry of $\frac{n_B}{s} \simeq \frac{\epsilon}{g^*} = 10^{-10}$. The most efficient was obtained to date using approximation (based on observation) $\Delta m^2_{sol} << \Delta m^2_{atm}$ [14]. It leads to

$$\begin{aligned} |\epsilon_1^\phi| &\leq \frac{\epsilon_{DI}^\phi}{2}\sqrt{1-\left[\frac{(1-a)\tilde{m}_1}{(m_3-m_1)}\right]^2}\sqrt{(1+a)^2-\left[\frac{(m_3+m_1)}{\tilde{m}_1}\right]^2}, \\ a &= 2\mathrm{Re}\left[\frac{m_1 m_3}{\tilde{m}_1^2}\right]^{1/3}\left[-1-i\sqrt{\frac{(m_1^2+m_3^2+\tilde{m}_1^2)^3}{27m_1^2m_3^2\tilde{m}_1^2}-1}\right]^{1/3} \end{aligned}$$

14.2 Gauge interactions and dilution of L

The asymmetry estimated in the previous section would only give the value of the final lepton (baryon number) in an ideal world, where on one hand, the canal considered is the only possible decay for the N particles, and on the other hand, the desintegration products are strictly "frozen", i.e. don't participate in any further interaction which could modify L, and thus B. Neither is the case.

While it is too often forgotten, I need to insist here on the gauge structure of the model. While the introduction of the right-handed neutrino's N_i may seem justified at the level of the Standard model as a way to generate (via the see-saw mechanism) a very small mass for their observed, mostly left-handed partners, things must be seen with a different eye when dealing with the high energy scales considered here. In this context indeed, closer to the grand unification scale than to the low energy domain, we must consider how such right-handed neutrinos enter the unification pattern. It seems a poor approach to merely add an unjustified singlet to the already unusual set of representations needed by $SU(5)$ to yield $1 + 5 + \overline{10}$. Instead, the attractive proposal is to consider $S0(10)$ keeping in mind that its smallest representation precisely decomposes under $SU(5)$ as : $16 = 1 + 5 + \overline{10}$.

We are not interested here in the details of the unification group, or in its breaking patterns (through $SU(5)$, Pati-Salam, or directly into the Standard model), but it is important to realize that the mechanism giving a (large) Majorana mass to the N is also at play in the symmetry breaking of the group, and will in general be associated *at least* to a set of gauge bosons transforming like $SU(2)_R$. We have all reason to expect (apart for fine-tuning) that the mass of such gauge bosons will be roughly in the same range as the mass of the N.

They offer then important new decay channels for the N neutrinos, either directly into W_R, or, if the latter happens to be too heavy, into light leptons.

The net lepton number generated in the decays (we will discuss rescattering and annihilation later) is thus diluted by these new decays. We call X this dilution factor.[15][16]

$$\epsilon_1^{tot} = \frac{\epsilon_1^{\phi}}{1 + X}$$

We will not discuss in details the various scenarios here. In short, it turns out that the presence of the very large 2-body decay channel quite generally

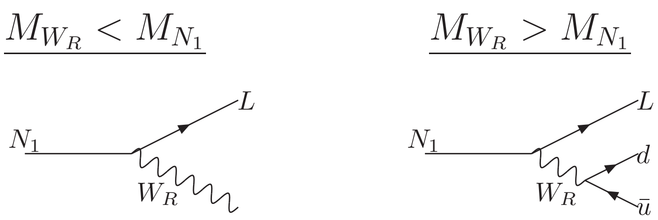

Figure 4: Additional relevant decay channels diluting the CP asymmetry according to whether $M_1 > M_{W_R}$ or $M_1 < M_{W_R}$.

induces too much dilution (typically 10^4 to 10^5) for leptogenesis to yield a sufficient baryon number of the Universe. We must thus require $M_{W_R} > M_1$. The 3-body decay rate then reads

$$\Gamma_1^{3b} = \frac{3g^4}{2^{10}\pi^3} \frac{M_1^5}{M_{W_R}^4}$$

and leads to a dilution factor

$$X = \frac{3g^4 v^2}{2^7 \pi^2} \frac{1}{\tilde{m}_1 M_1 a_R^2}$$

where $a_R = M_{W_R}^2 / M_1^2$.

Before closing this section, we must however keep in mind that the presence of the gauge couplings is not necessarily unfavourable to the lepto- or baryogenesis scheme. We will see below indeed that, apart from the dilution they bring inevitably, gauge interactions may play an important role in reconstituting the N population during re-heating after inflation. In this way, they in fact help evade a lower bound on neutrino masses!

14.3 Rescattering, diffusion

Once again, if the Universe would cool very rapidly just after lepton number generation, this section would we useless. In a realistic cosmological scheme, we must however take into account a number of reactions, mostly active at temperatures close to the N_1 decoupling, which can wash out all or part of the expected lepton number. The simplest way in this review to go through this part is to give a list of Feynman diagrams contributing to this process. This part is usually dealt with in an approximate scheme, using Boltzman

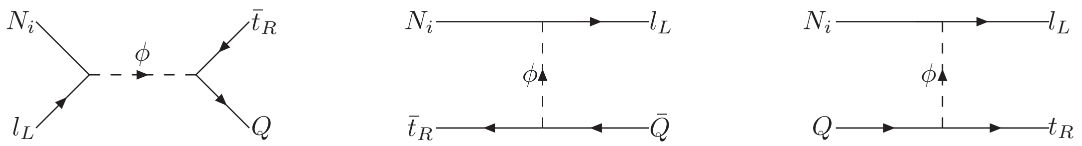

Figure 5: $\Delta L = 1$ diffusion interactions.

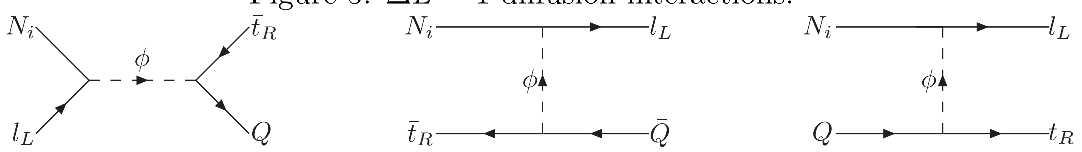

Figure 6: $\Delta L = 2$ diffusion interactions.

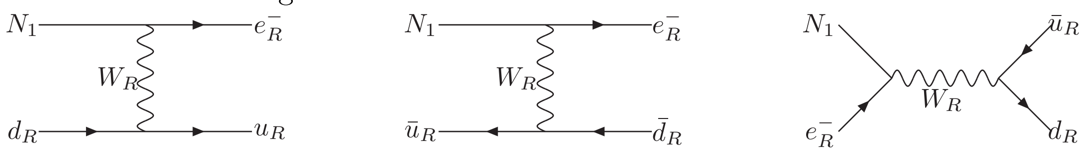

Figure 7: Diffusion interactions with one N_1

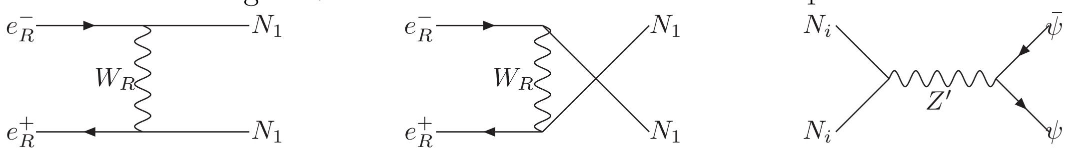

Figure 8: Diffusion interactions with two N_1

evolution equations (one delicate point is to avoid double counting with the real decay processes already considered).

All of these effects (which require detailed calculations) are usually lumped into one "efficiency" parameter, η_{eff}, leading to the global formula (remember that ϵ_1 already includes the gauge dilution):

$$\eta_{eff} = \frac{Y_L(z=\infty)}{\epsilon_1^{\phi} Y_{N_1}^{eq}(init.)}$$

where Y refers to the abundance of particles.

14.4 Thermal and re-heating scenarios

Even accepting the general scheme of leptogenesis, followed by conversion of L to B at the electroweak transition, we must still specify one important point about the cosmological model.

This far, we have worked in the hypothesis of "thermal" leptogenesis, namely we have considered a Universe which is initially hot, where all particles reach their equilibrium abundance, and a subsequent cooling down, during which some heavy particle (not able to decay fast enough to match the cooling) become "relics", which then decay out-of-equilibrium.

For a number of reasons (the most direct one is the near-isotropy of the fossil radiation, which is difficult to explain in a thermal Universe, some parts of which have not come in causal contact in the above picture), astrophysicists now favour "inflation" scenarios, which are phases of rapid expansion of the Universe. Typically such inflation is controlled by the evolution of a scalar field (to which we will refer as "inflaton" in a generic way). In such a scheme, the initial distribution of particles is vastly diluted, and becomes negligeable. The new distribution of particles after inflation is generated typically by the fluctuations of the inflaton field.

Of course, all depends on the way this field couples to matter. Actually, the coupling to N could even be favoured, but the assumption retained below is that the N particles are not created directly by the inflation, but that their population must be re-built through the N interactions with other matter. Here the gauge coupling may come to the rescue, as we see in fig 9.

In this figure indeed, we have distinguished the thermal (full lines) and re-heating cases (dashed lines) for a few values of the W_R mass, and shown the corresponding iso-dilution curves.

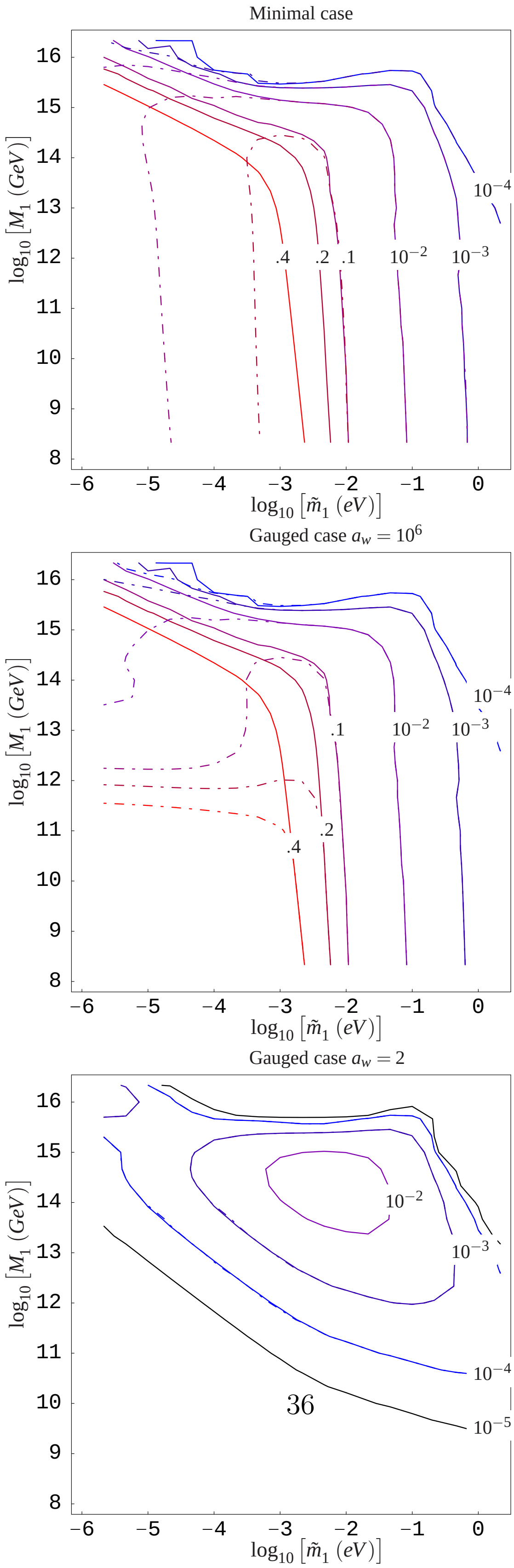

Figure 9: dilution factors as a function of the $\tilde{m}_1$ and M_1, for various ratios of Majorana and Gauge masses. The continuous lines refer to the thermal case, the dashed ones to the re-heating situation

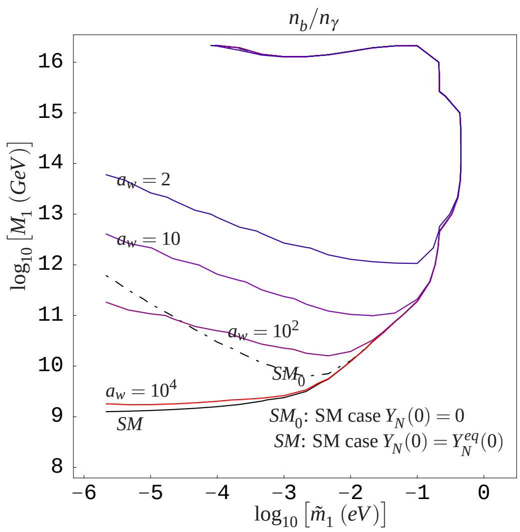

Figure 10: baryon number as a function of M_1 and $\tilde{m}_1$, for various values of M_{W_R}

The effect is striking: for very heavy M_{W_R}, a small value of $\tilde{m}_1$ is forbidden in the case of reheating: this is easily understood, as this small value means that the N_1 particle is virtually uncoupled to the light fermions, and thus its population cannot be rebuilt. However, even a modest W_R , corresponding to the graph where $M_{W_R}/m_{N_1} = 10^3$ is sufficient to drop this constraint. In other words, it turns out that even a small effect of heavy gauge bosons eliminates a potential lower bound on neutrino masses at low energy. The graph with a lighter M_{W_R}shows both this effect (the dash-dotted line is completely confused here with the plain one), and the larger dilution effect, as read from the curves.

14.5 Conclusion on Leptogenesis

We have briefly sketched above the main steps leading to a calculation of leptogenesis. From the orders of magnitude, it comes clearly that leptogenesis (followed by lepton conversion to baryon at the electroweak scale) is a strong contender to explain the baryon number of the Universe.

We show in fig 10 contour plots in the $M_1 - \tilde{m}_1$ plane for specific values of M_{W_R}, showing that a confortable space of parameters is allowed (once again, dashed lines refer to the reheating case)

Many detailed models exist, and they try to link the observed baryon number to the value of the quark masses. We can only refer the reader to the current litterature for this, in particular [9] [14], but we want to stress here that such a step is necessarily very speculative, and we hope to have shown that important effects should not be neglected (a fairly obvious, and dramatic one is that the gauge interactions naturally associated to the heavy neutrinos cannot be neglected).

15 Conclusion

The series of lectures summarized here only aimed at presenting an hopefully pedagogical introduction to the field of baryogenesis, including its currently most favoured approach, leptogenesis. As is plainly obvious, such a tentative is a real challenge, since the subject had to be presented to physicists and astrophysicists from very different backgrounds, and, on the other hand, the number of concepts (even in particle physics alone) brought into play is extraordinarily large.

16 Acknowledgements

A first series of lectures (rather more general) was given at the "Ecole de Gif" hold in Strasbourg in 2002. These noted were updated for a more topical presentation at the ITEP school 2005. I wish to thank the organisers of both meetings (and in particular Daniel Bloch, Misha Vysotsky, and Michael Danilov) for their invitation and constant encouragements. On a more formal side, I want to thank the IISN (Communauté française de Belgique) and the Belgian Federal Science Policy Office (under IAP V/27 Fundamental Interactions) for their funding.

References

[1] It is impossible in the framework of this pedagogical introduction to give a comprehensive bibliography of what has become a very vast field. With the exception of a few specific references below, we refer thus the reader to the two following excellent and extensive reviews, and their

respective bibliographies:
A. Riotto and M. Trodden, *Recent progress in baryogenesys* hep-ph/9901362, Ann.Rev.Nucl.Part.Sci.**49**:35-75,1999
W. Büchmuller, R. D. Peccei and T. Yanagida, *Leptogenesis as the origin of matter*, hep-ph/0502169 **and references quoted**

[2] J.-M. Frère, "CP violation: From kaons to the universe," Surveys High Energ. Phys. **9** (1996) 203.

[3] A.D.Sakharov, JETP Lett. **5**,24 (1967).

[4] M.Yoshimura, Phys. Rev. Lett. **41**, 281 (1978).

[5] I. Affleck and M. Dine, Nucl. Phys. **B249**, 361 (1985)

[6] R. F. Klinkhammer and N. S. Manton, Phys. Rev. **D30**,2212,1996

[7] C. Jarlskog Z. Phys **C29**, 491, 1985

[8] M. Fukugita and T. Yanagida, Phys Lett.**B174**,45, 1986

[9] W Buchmuller and M. Plumacher, *Neutrino masses and the Baryon asymmetry*, hep-ph/0007176, Int.J.Mod.Phys.**A15**:5047-5086,2000,
W. Büchmuller, P. Di Bari, M. Plumacher, Nucl.Phys B665:445,2003

[10] S. Yu Khlebnikov and M. E. Shaposhnikov, Nucl. Phys. **B503**, 24, 1997

[11] J.-M. Frère, J.A.M. Vermaseren,M.B. Gavela, Phys.Lett **B103**:129,1983

[12] L.Covi, E. Roulet, F. Vissani, Phys. Lett. **B424**,101,(1998)
J.-M. Frère, F.-S. Ling, M.H.G. Tytgat, V. van Elewyck, *Leptogenesis with virtual Majorana neutrinos*, hep-ph/9901337, Phys.Rev.**D60**:016005,1999

[13] S. Davidson and A. Ibarra, Phys.Lett. **B535**,25 (2002)

[14] T. Hambye, Y. Lin, A. Notari, M. Papucci and A. Strumia, hep-ph/0312203 Nucl.Phys **B695**:169,2004

[15] S. Carlier, J.-M. Frère, F.-S. Ling *Gauge dilution and Leptogenesis* ,hep-ph/9903300,Phys.Rev.**D60**:096003,1999

[16] N.Cosme, JHEP **0408**:027,2004