Revisiting an old friend in the light of Adiabatic Quantum Computing

Steve Abel *(IPPP)*

Recent papers mainly based on work w/ Juan Craido and Michael Spannowsky

Oxford ‘92

Oxford ‘92

Oxford '92

Life:

Reputation as a forceful character !

Life:

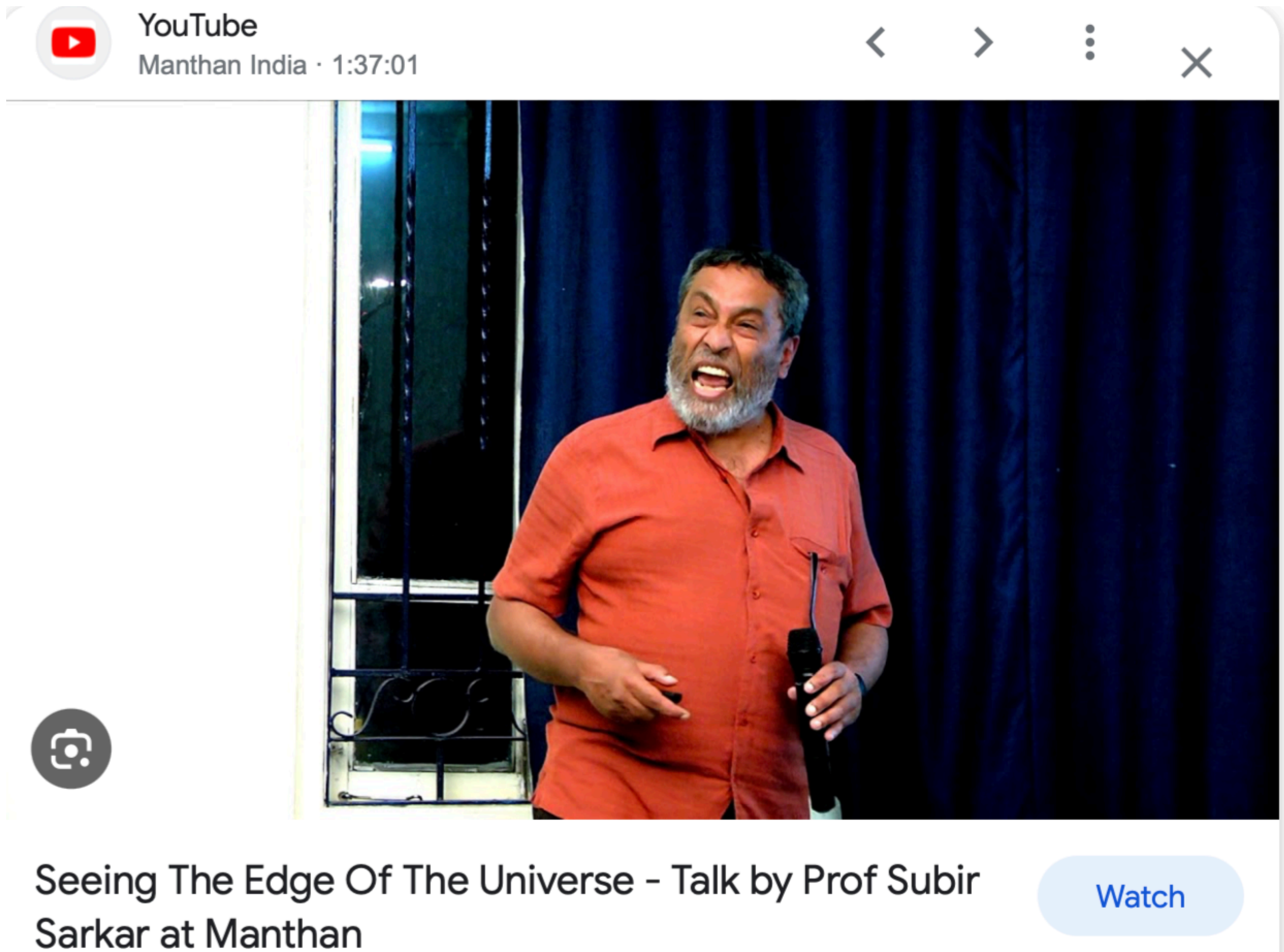
Reputation as a forceful character !

However after watching a couple of talks you realise he’s actually very mild …

Life:

Reputation as a forceful character !

However after watching a couple of talks you realise he's actually very mild …

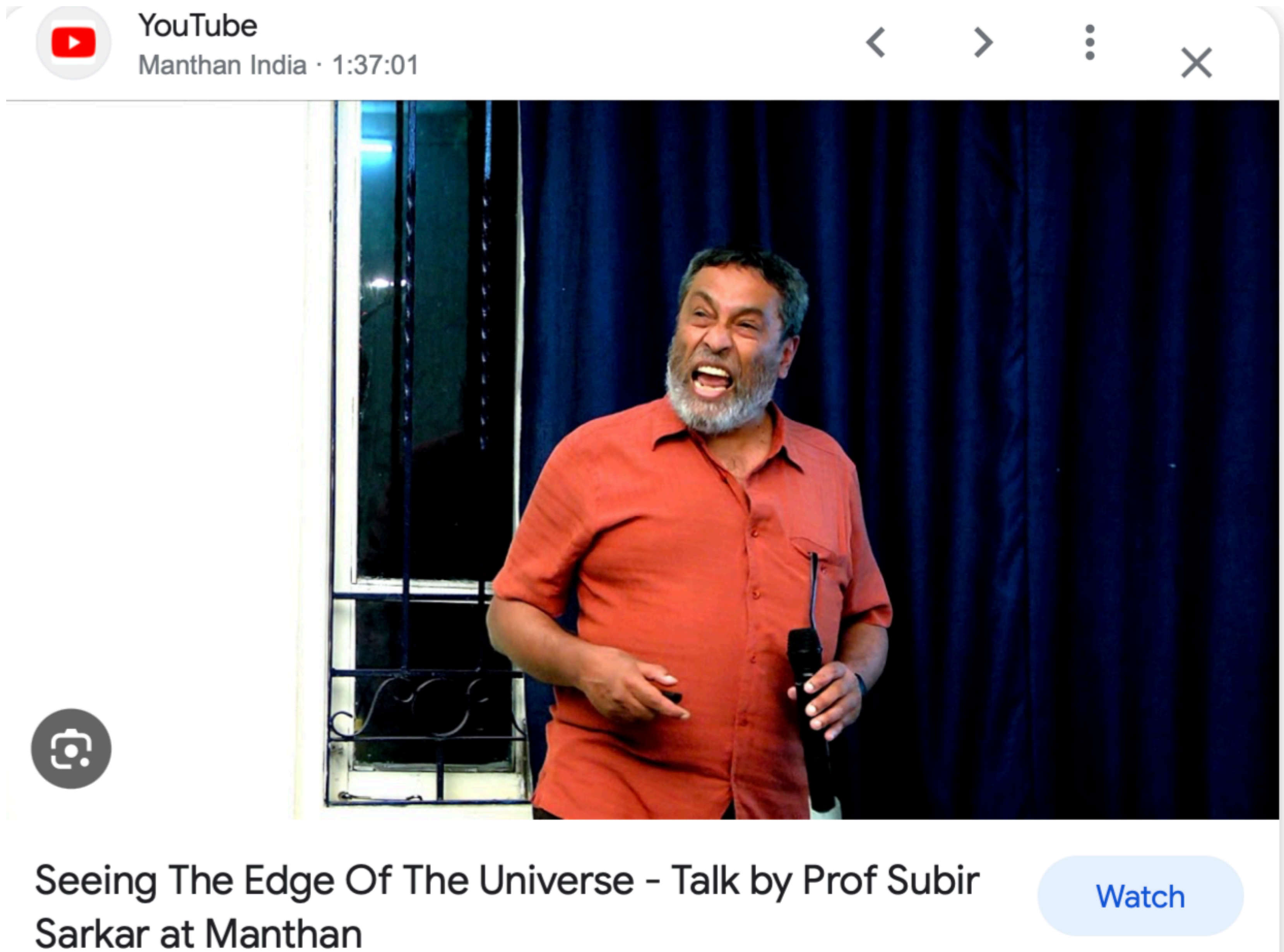


Life:

Reputation as a forceful character !

However after watching a couple of talks you realise he’s actually very mild …

One of the voices in my head about integrity in science and life.



Papers:

Nuclear Physics B392 (1993) 83–110
North-Holland

NUCLEAR
PHYSICS B

Neutralino dark matter in a class of unified theories

S.A. Abel and S. Sarkar

Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK

I.B. Whittingham

Department of Physics, James Cook University, Townsville, Australia 4811

Received 16 June 1992
(Revised 2 November 1992)
Accepted for publication 3 December 1992

Papers:

ELSEVIER

12 January 1995

Physics Letters B 342 (1995) 40–46

PHYSICS LETTERS B

Cosmological constraints on perturbative supersymmetry breaking

S.A. Abel [a], S. Sarkar [b,1]

[a] *Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK*

[b] *Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK*

Received 20 September 1994
Editor: P.V. Landshoff

Abstract

We discuss the cosmology of string models with perturbative supersymmetry breaking at a scale of $\mathcal{O}(\text{TeV})$. Such models exhibit Kaluza-Klein like spectra and contain unstable massive gravitinos/gravitons. We find that considerations of primordial nucleosynthesis constrain the maximum temperature following inflation to be not much larger than the supersymmetry breaking scale. This imposes conflicting requirements on the scalar field driving inflation, making it rather difficult to construct a consistent cosmological history for such models.

Papers:

ELSEVIER

Nuclear Physics B 454 (1995) 663–681

NUCLEAR
PHYSICS B

On the cosmological domain wall problem for the minimally extended supersymmetric standard model

S.A. Abel [a], S. Sarkar [b,1], P.L. White [b]

[a] *Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK*
[b] *Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK*

Received 28 June 1995; accepted 11 September 1995

Papers:

Next-to-Minimal Supersymmetric Standard Model (NMSSM) has Z3 symmetry: Good arguments to suppose such a global symmetry is broken by gravity …

$$\varepsilon \sim \lambda' \sigma M_W^2 / M_{Pl}$$

Remove domain walls before BBN :

$$\lambda' \gtrsim 10^{-7}$$

But to avoid destabilisation of EW scale require :

$$\lambda' \lesssim 3 \times 10^{-11}$$

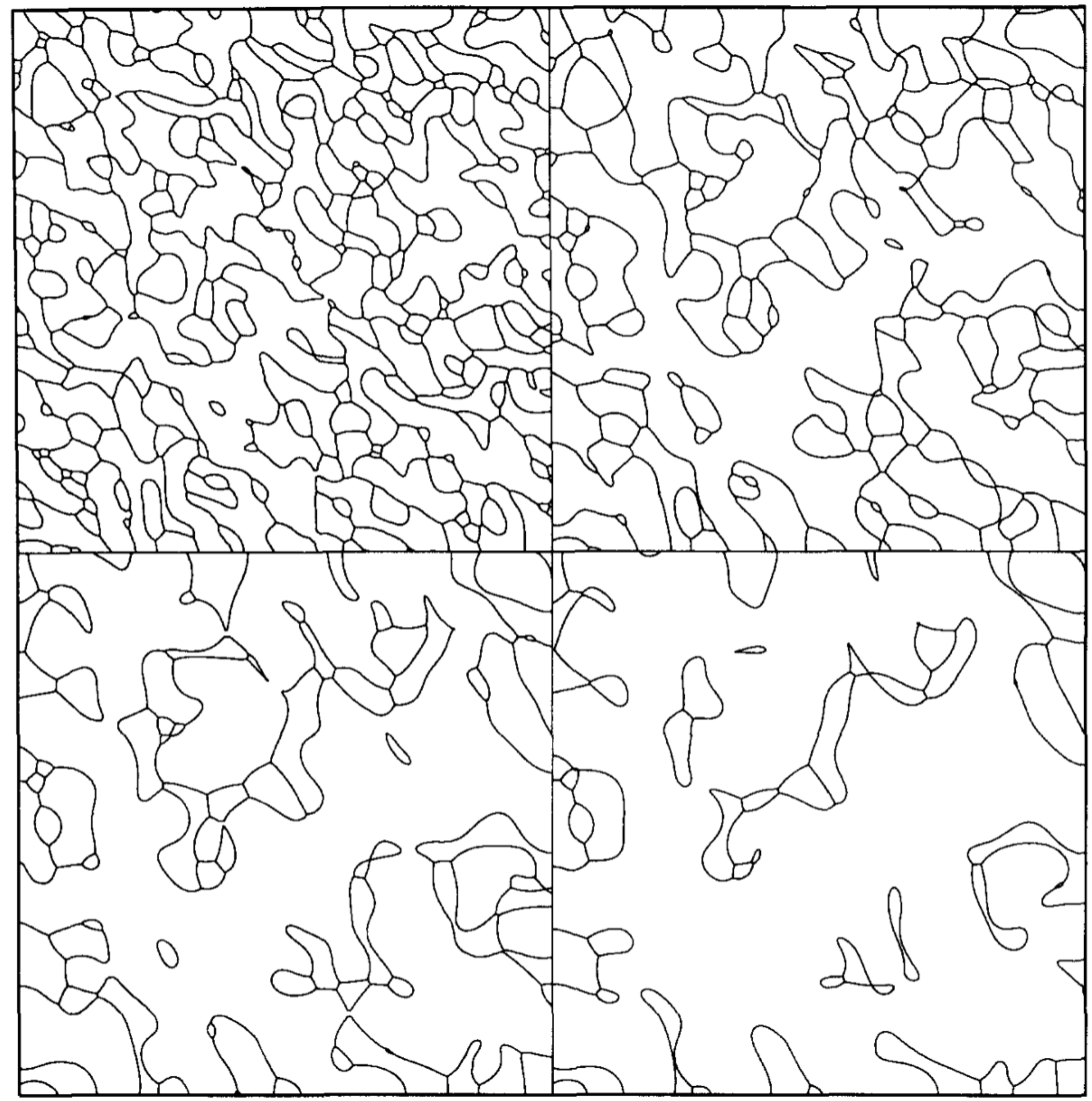


Fig. 3. A typical example of the evolution of the wall network with a pressure term of order $\sigma M_W^2 / M_{Pl}$. The figure shows the wall network at four epochs separated by an interval of 10^{-10} sec, beginning at the time when pressure starts to dominate the evolution.

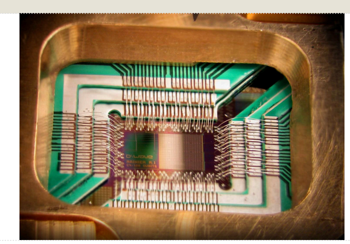
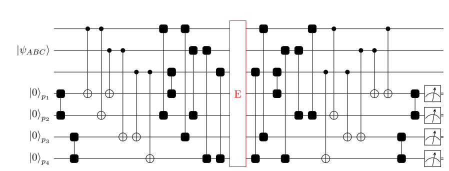
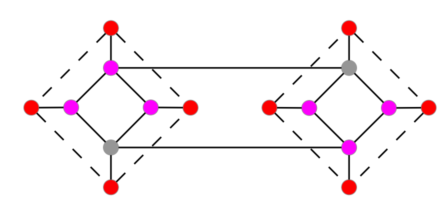
Papers:

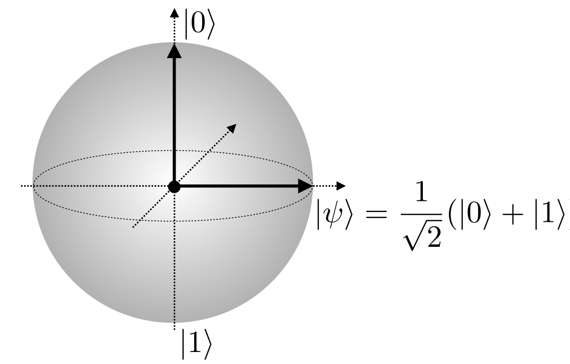
Possible get-out clauses:

1) Z3 is anomalous w.r.t. SU(3)
2) Giudice Masiero generated mu -term (needs R-symmetry or similar, SAA 1996)
3) Z3 symmetry broken at high scale, $M_{\mathtt{contrived}}$, in the visible sector.
4) Z3 is gauged discrete symmetry (Lazarides-Shafi mechanism). Begin with network of cosmic strings: after EWSB joined by network of domain walls which makes them collapse — also in principle walls able to decay by forming a hole with a cosmic-string boundary in them.
5) Initial biasing of distributions of vacua (Coulson, Lalak and Ovrut 1995, Larsson, Sarkar, White 1997) — e.g. symmetry breaking occurs through some intermediate phase which allows one minimum to be preferentially populated.

Adiabatic Quantum Computing for Neural Networks

Background: Quantum computing has a long and distinguished history but is only now becoming practicable. (Feynman ’81, Zalka '96, Jordan, Lee, Preskill … see Preskill 1811.10085 for review). Two types of Quantum Computer:

Type	Discrete Gate	Dedicated Quantum Annealer
Property	Universal (any quantum algorithm can be expressed)	Not universal — certain quantum systems
How?	IBM - Qiskit ~127 Qubits	DWave - LEAP ~7000 Qubits
What?		
		



- Both types operate on the Bloch sphere: basically measuring $\sigma_i^Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ where $(\sigma^Z|0\rangle = |0\rangle,\, \sigma^Z|1\rangle = -|1\rangle)$ are the possible eigenvector eqns

- Each *i* represents a single qubit

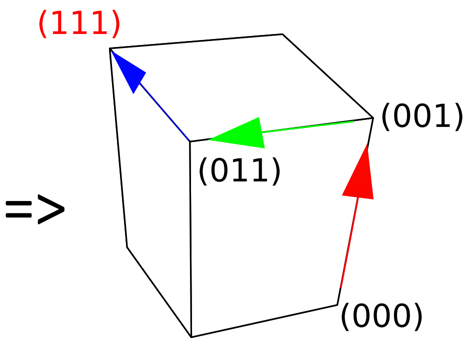
- A discrete quantum gate system is good for looking at things like entanglement, Bell's inequality etc. Also discrete problems, cryptographical problems, Shor's, Grover's algorithms, etc.

- Quantum annealing is good for looking at network optimisation problems. In practice often based on the general transverse field Ising model (Appolloni, Cesa-Bianchi, de Falco (1988), Kadowaki, Nishimori):

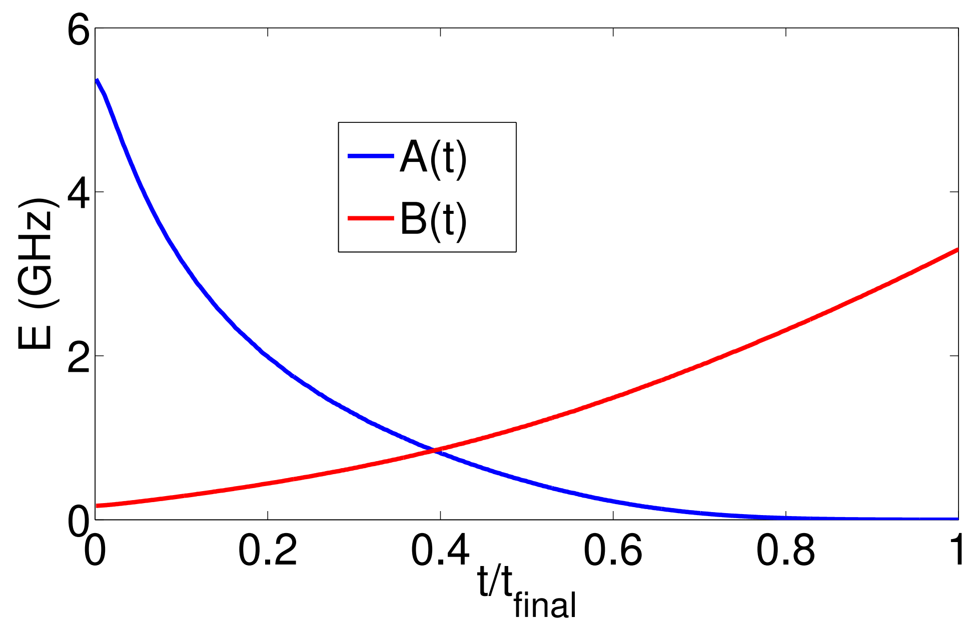
- What does the “anneal” mean?

$$\mathcal{H}(t) = A(t)H_0 + B(t)H_1$$
$$= A(t)\sum_{\ell}\sigma_{\ell}^{X} + B(t)\left(\sum_{\ell}h_{\ell}\sigma_{\ell}^{Z} + \sum_{\ell m}J_{\ell m}\sigma_{\ell}^{Z}\sigma_{m}^{Z}\right)$$

A(t) > 0 induces bit-hopping in the Hamming/Hilbert space ==>



The original idea is to start in the groundstate of the simple H_0 and dial the parameters to land in the global minimum (i.e. the solution) of some “problem Hamiltonian” described by H_1

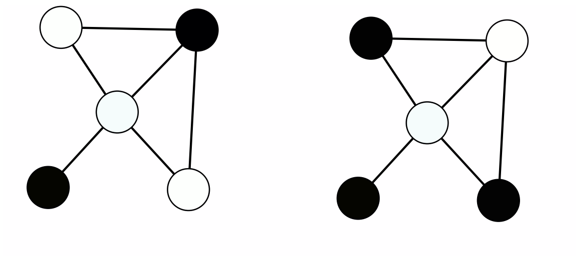


A(t) and *B(t)* is called the anneal schedule
i.e. we take *A(0)=B(1)=1* and *A(1)=B(0)=0.*

- ***Adiabatic Quantum Computing*** (AQC) means to strictly stay in the groundstate at all times (Farhi, Goldstone, Gutmann, Lapan, Lundgren, Preda, 2000)

How do we use it? Encoding network problems in a general Ising model

- Example: maximum number of coloured vertices on a graph so that none touch? NP-hard problem.



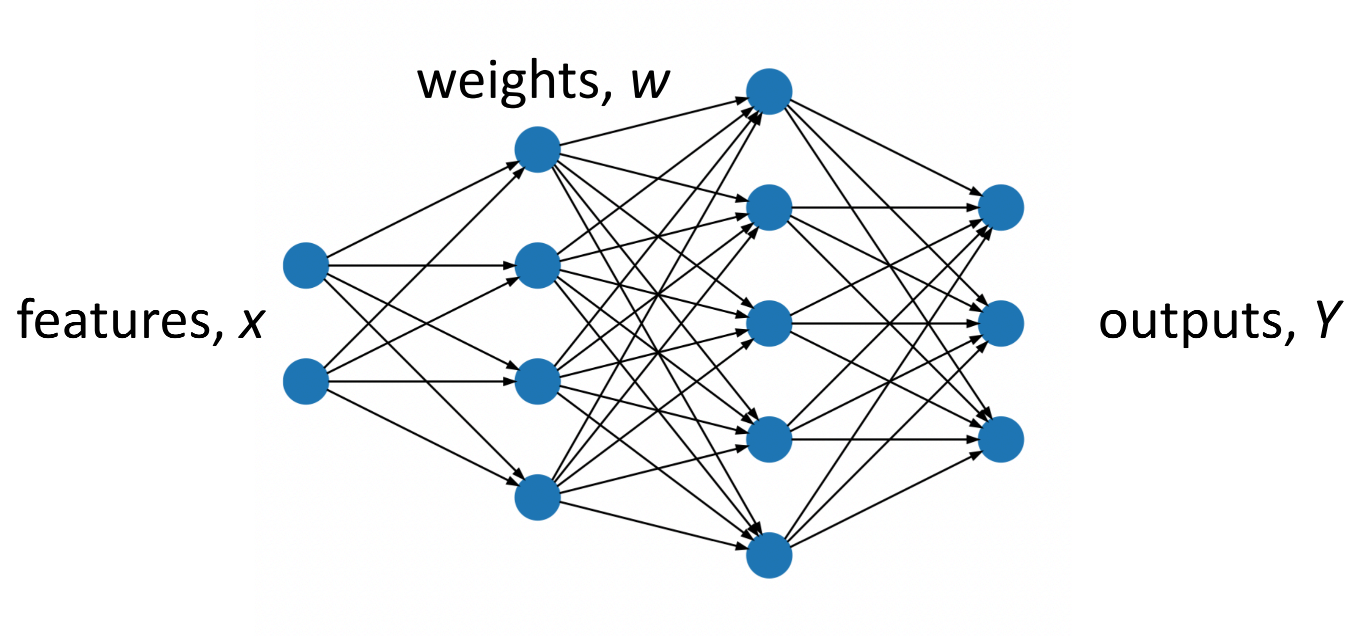
- Let non-coloured vertices have $\sigma_i^Z = -1$ and coloured ones have $\sigma_i^Z = +1$.

- Add a reward for every coloured vertex, and for each link between vertices *i,j* we add a penalty if there are two +1 eigenvalues:

$$H_1 = -\Lambda \sum_i \sigma_i^Z + \sum_{\text{linked pairs } \{i,j\}} \left[\sigma_i^Z + \sigma_j^Z + \sigma_i^Z \sigma_j^Z\right]$$

Application: Completely Quantum Neural Networks

Recap of classical NNs: the AI in your phone consists of a NN that encodes the solution to a class of problems in weights and biases:



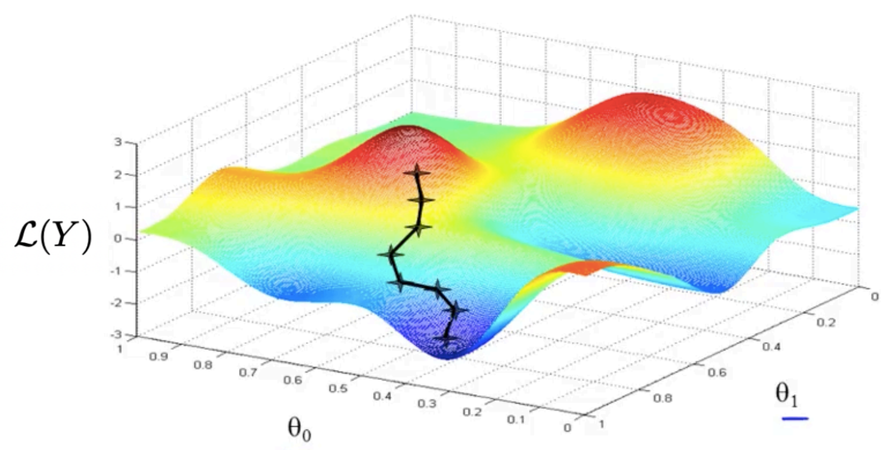
NN produces outputs *Y* by passing inputs *x* through layers with activation functions *g* as follows*:*

$$L_i(x) = g\left(\sum_j w_{ij}x_i + b_i\right) \qquad Y = L^{(n)} \circ \ldots \circ L^{(0)}$$

To make the network learn (in a supervised way), we define a loss function that we minimise for a whole load of previous data to determine all the weights and biases (e.g. for classification with data labelled data by *a*):

$$\mathcal{L}(Y) = \frac{1}{N_d}\sum_a |y_a - Y(x_a)|^2$$

Classically: minimise $\mathcal{L}$ using gradient descent and backpropagation



The loss function establishes a hypersurface for which we can try to find a minimum usually using gradient descent

Gradient descent for every weight $w_{ij}^{(l)}$ and every bias $b_i^{(l)}$ in the NN looks like:

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha \frac{\partial}{\partial w_{ij}^{(l)}} \mathcal{L}(w,b)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} \mathcal{L}(w,b)$$

in short: $w_{new} = w_{old} - \alpha * \nabla error$

where α is the learning rate

Difficulty training NNs: When the NN is small (and efficient) the training process can be difficult. Also discrete or binary networks (weights = 0 or 1) are very hard to train as gradient descent doesn't work. A summary of the problems:

- Badly conditioned curvature (ravines)
- Local minima
- Weight degeneracy (symmetries in weights)
- Dead and saturated weights (plates in the loss-function landscape)

Quantum training of NNs: The training process can be one of the lengthiest parts of the process: can we use a quantum annealer to train? (After all it is built to minimise loss functions.)

How best to do this?

If we think about $\mathcal{L}(Y) = \frac{1}{N_d}\sum_a |y_a - Y(x_a)|^2$ we want to avoid having to encode each data point in qubits

We can instead encode the weights and biases in qubits in binary fashion and read off their values.

Examples using Quantum Annealer of D-wave: we took a *single* hidden layer:

$$Y_{v,w}(x_j) \;=\; v_i g(w_{ij} x_j) + v_0$$

The activation function must be nonlinear for a NN to work, but it can be simple: $g(x) = (1+x)^2/4$

Then what appears in the loss function $\mathcal{L}(Y) \;=\; \frac{1}{N_d} \sum_a |y_a - Y(x_a)|^2$ is

$$\begin{aligned} Y_{v,w}(x) \;&=\; y_a - \tfrac{1}{4} - \tfrac{1}{2} v_i w_{ij} x_{aj} \\ &\quad - \tfrac{1}{4} v_i w_{ij} w_{ij'} x_{aj} x_{aj'} \end{aligned}$$

with the weights being encoded as fractional binaries …

$$\omega \;=\; -1 + \frac{1}{1-2^{-\beta}} \sum_{\alpha=0}^{\beta} 2^{-\alpha} \tau_{\alpha}^{\omega}$$

e.g. 2D datasets = “circles”, “quadrants”, “bands” and t-tbar yields a classification curve. (The features for the latter are the highest transverse momentum of a b-jet and the missing energy, in simulated LHC pp collisions.)

SAA, Criado, Spannowsky

Advantage: our weights and biases are all discretised due to the “qubitisation”. A standard NN cannot be trained very well for discrete weights and biases.

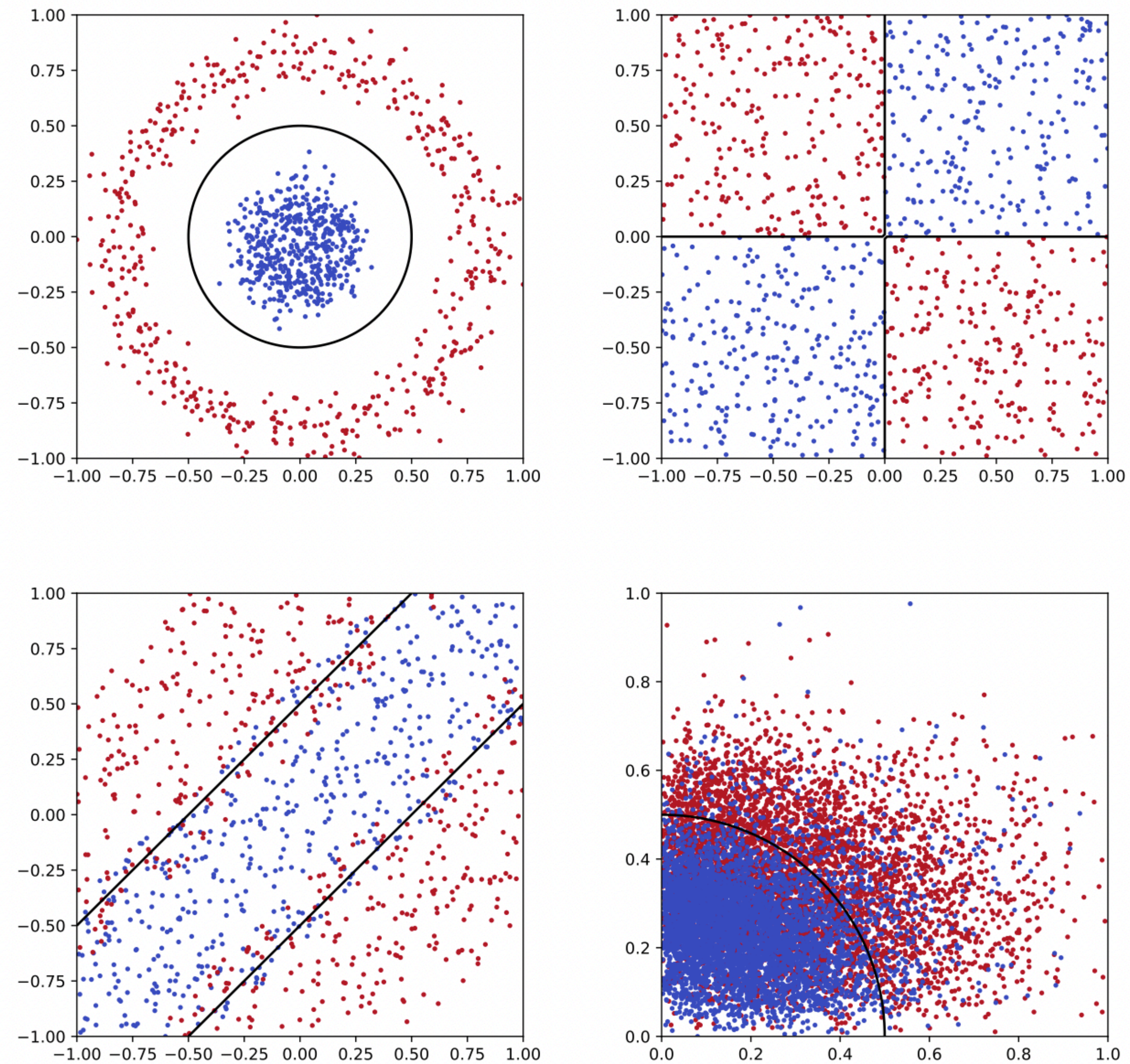
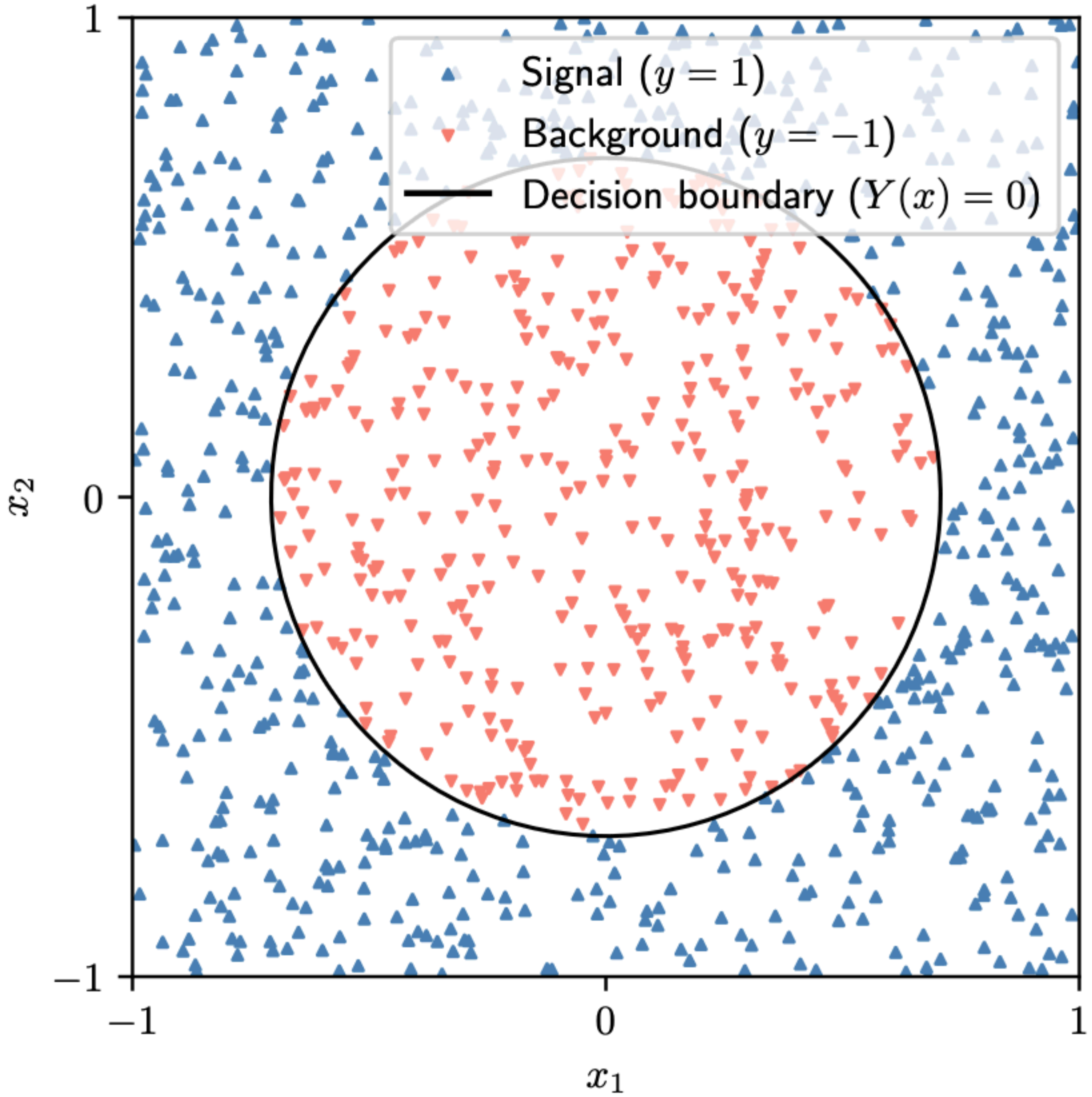


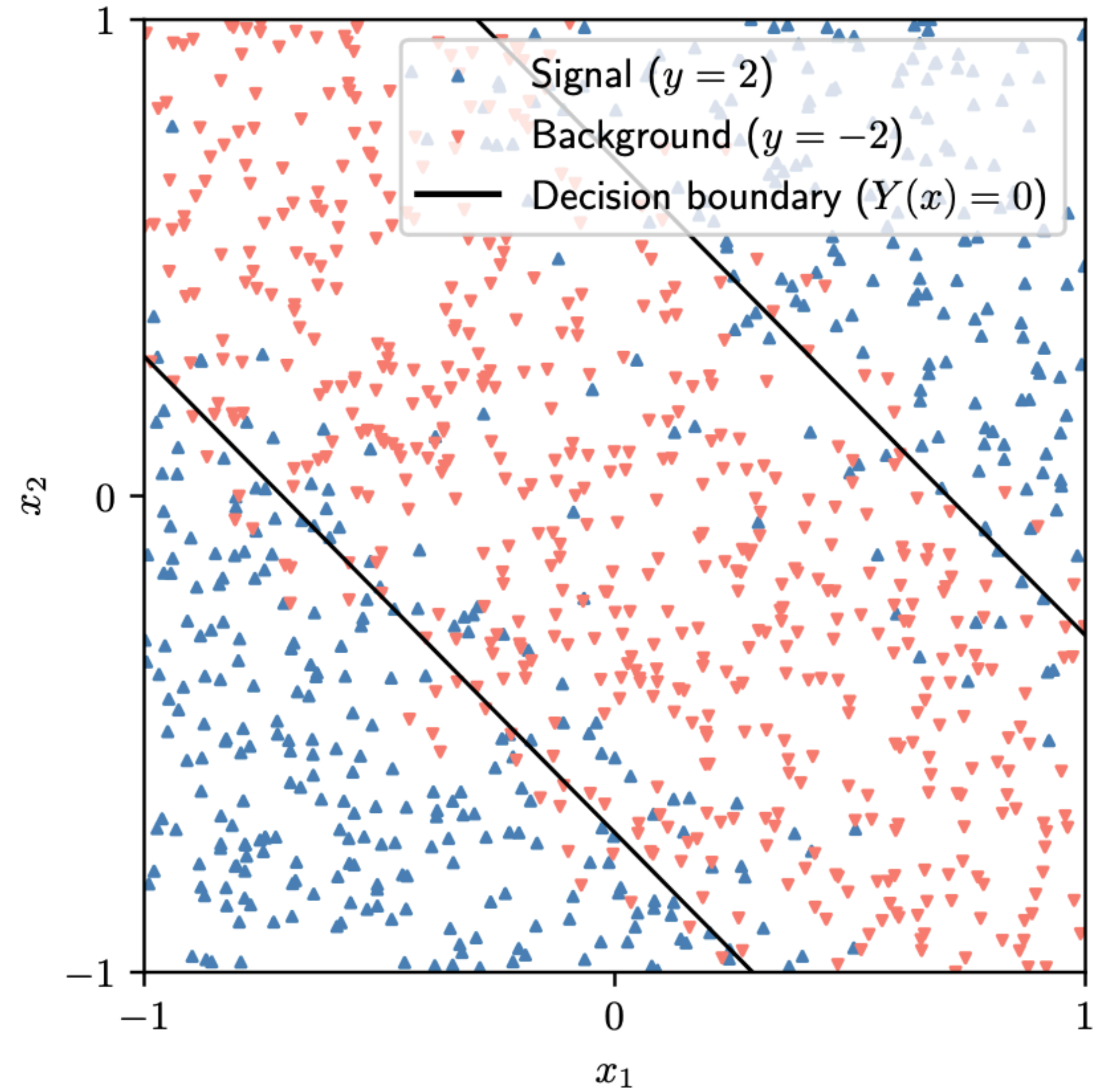
Figure 4: Decision boundary obtained with the quantum NN for each dataset.

More recently using strict AQC on gate quantum computers using Qibo (Bravo-Prieto, Carrazza, Efthymiou, Garcıa-Martin, Garcia-Saez, Latorre, Ramos-Calderer, Perez-Salinas, https://qibo.science/)

Unlike D-wave gate quantum computers are universal so much more flexible:

SAA, Criado, Spannowsky





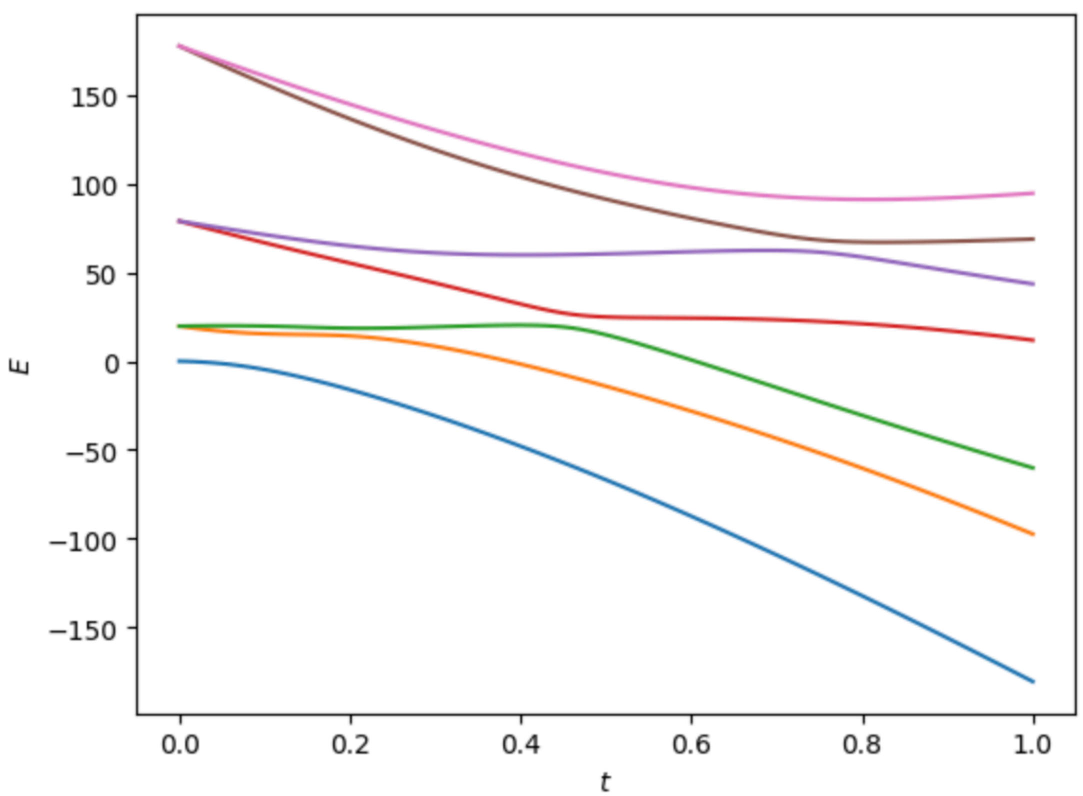
More on Quantum Annealers versus Adiabatic Quantum Computing: domain walls revisited

More on Quantum Annealers versus Adiabatic Quantum Computing:

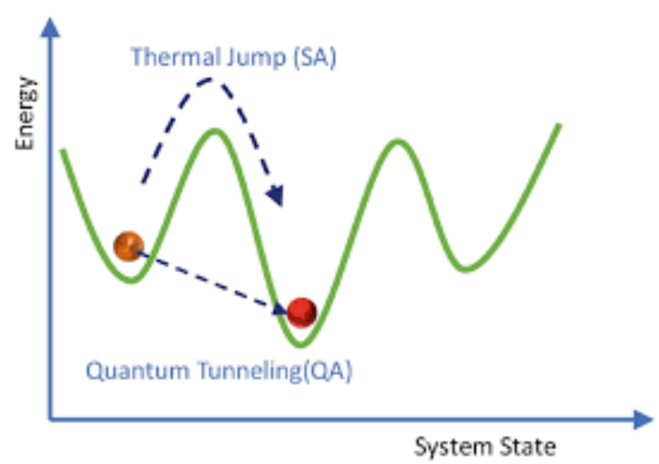
Quantum annealers like D-wave's are diabatic and dissipative — they lose energy (and coherence) but are great for finding ground states by tunnelling.

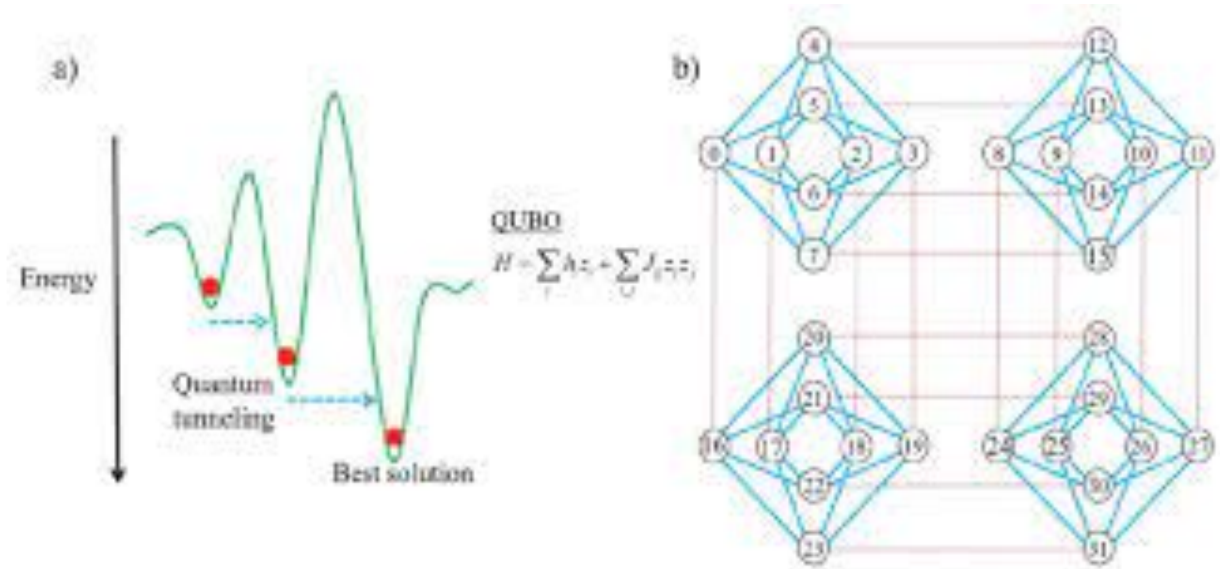
However the idea of AQC is to remain in the ground state as we adjust the Hamiltonian adiabatically to end up in the difficult Hamiltonian. (Farhi, Goldstone, Gutmann, Sipser)

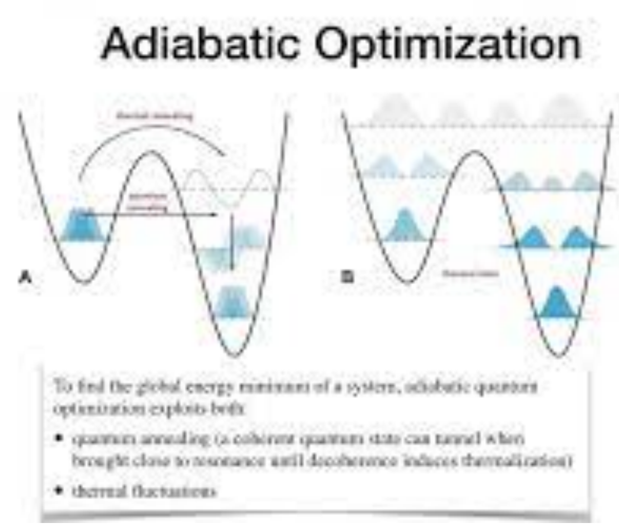
i.e. The evolution of the spectrum should look something like …

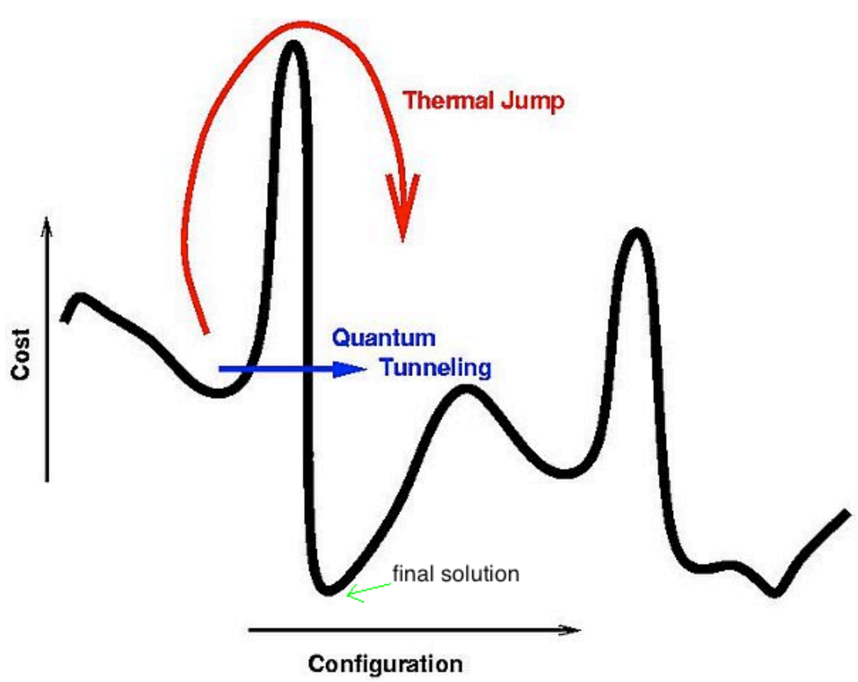


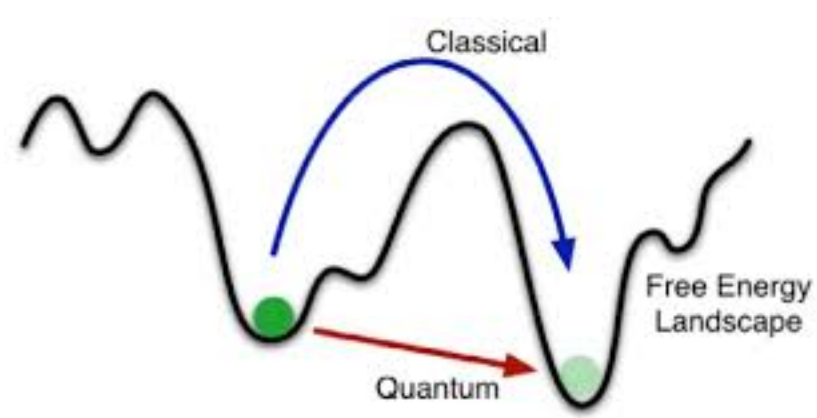
What is an Adiabatic Quantum Computer really doing? (Not what the internet thinks)





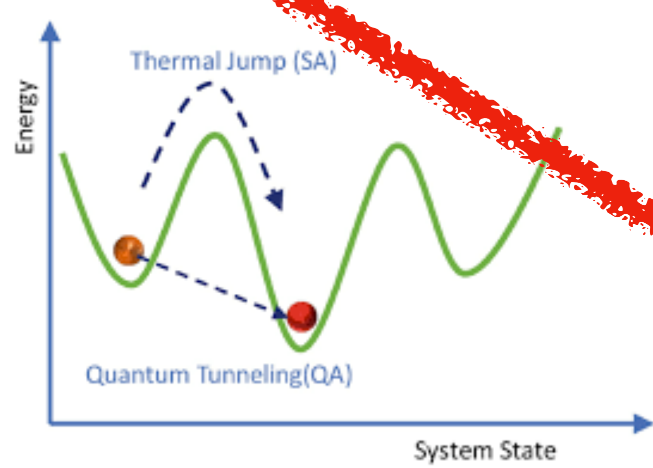


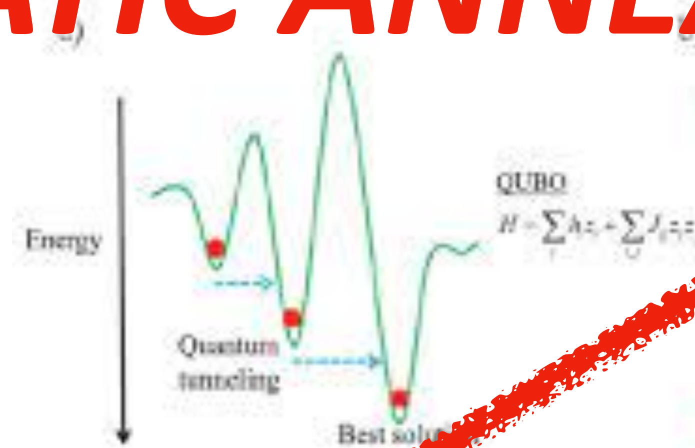


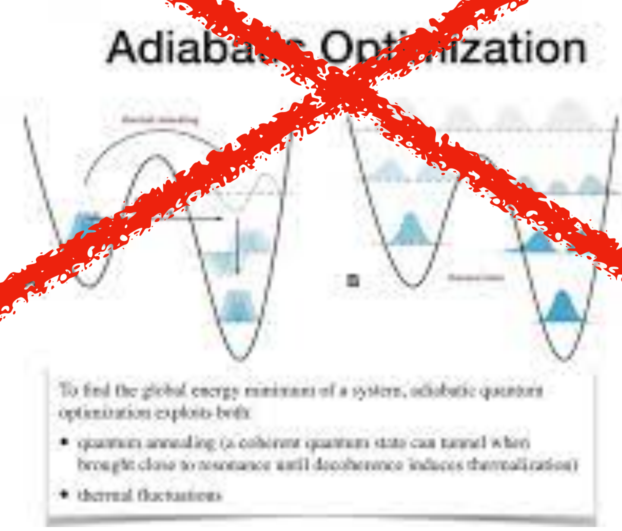


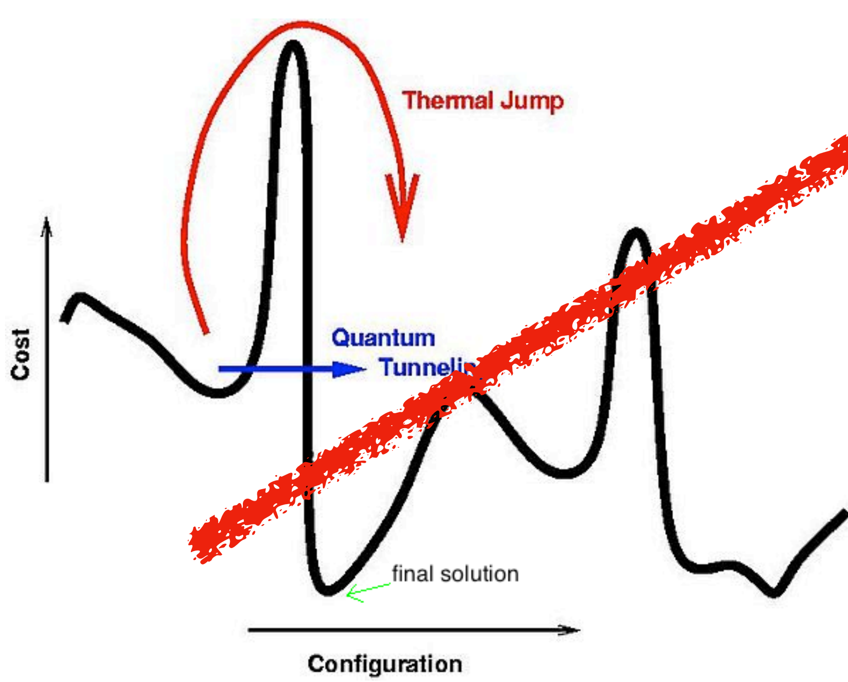
What is an Adiabatic Quantum Computer really doing? (Not what the internet thinks)

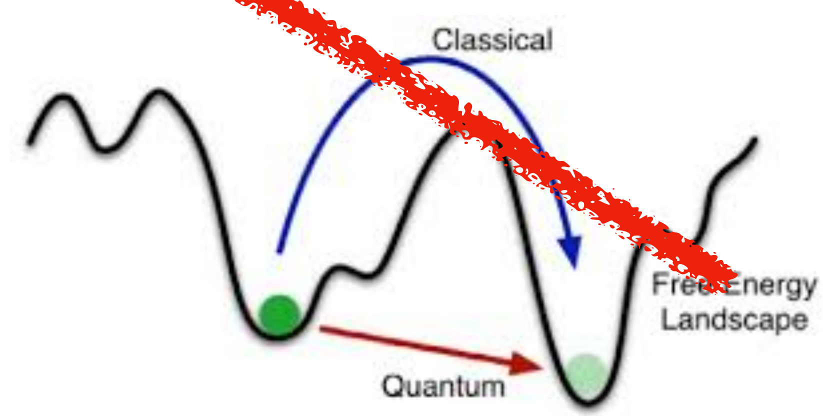
THIS IS DIABATIC ANNEALING









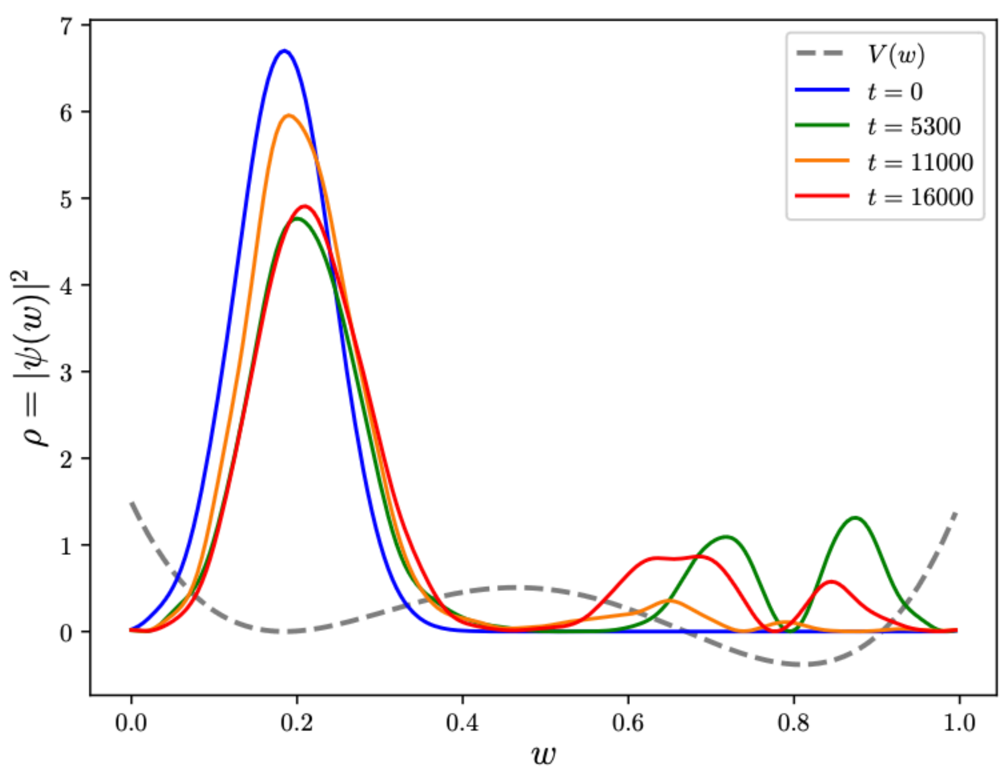


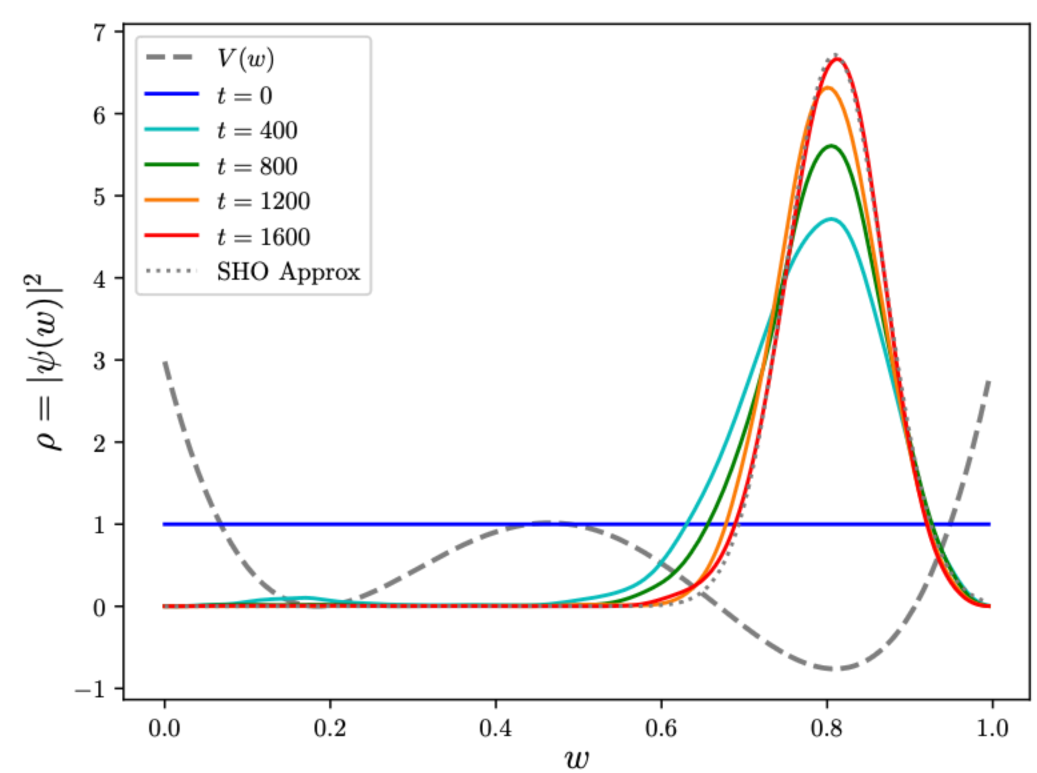
What is an Adiabatic Quantum Computer really doing? (Not what the internet thinks)

In AQC the coupling A*(t)* doesn't really induce bit-hopping in the Hamming/Hilbert space but it is taking the ground state at all times and therefore sampling the entire space at all times. There isn't really any "tunnelling" because the system is never stuck. To make it clear let's look at minimising a simple 1D potential which *grows adiabatically* - i.e.:

$$H_0 = p^2/2m \quad H_1 = p^2/2m + V(w)$$

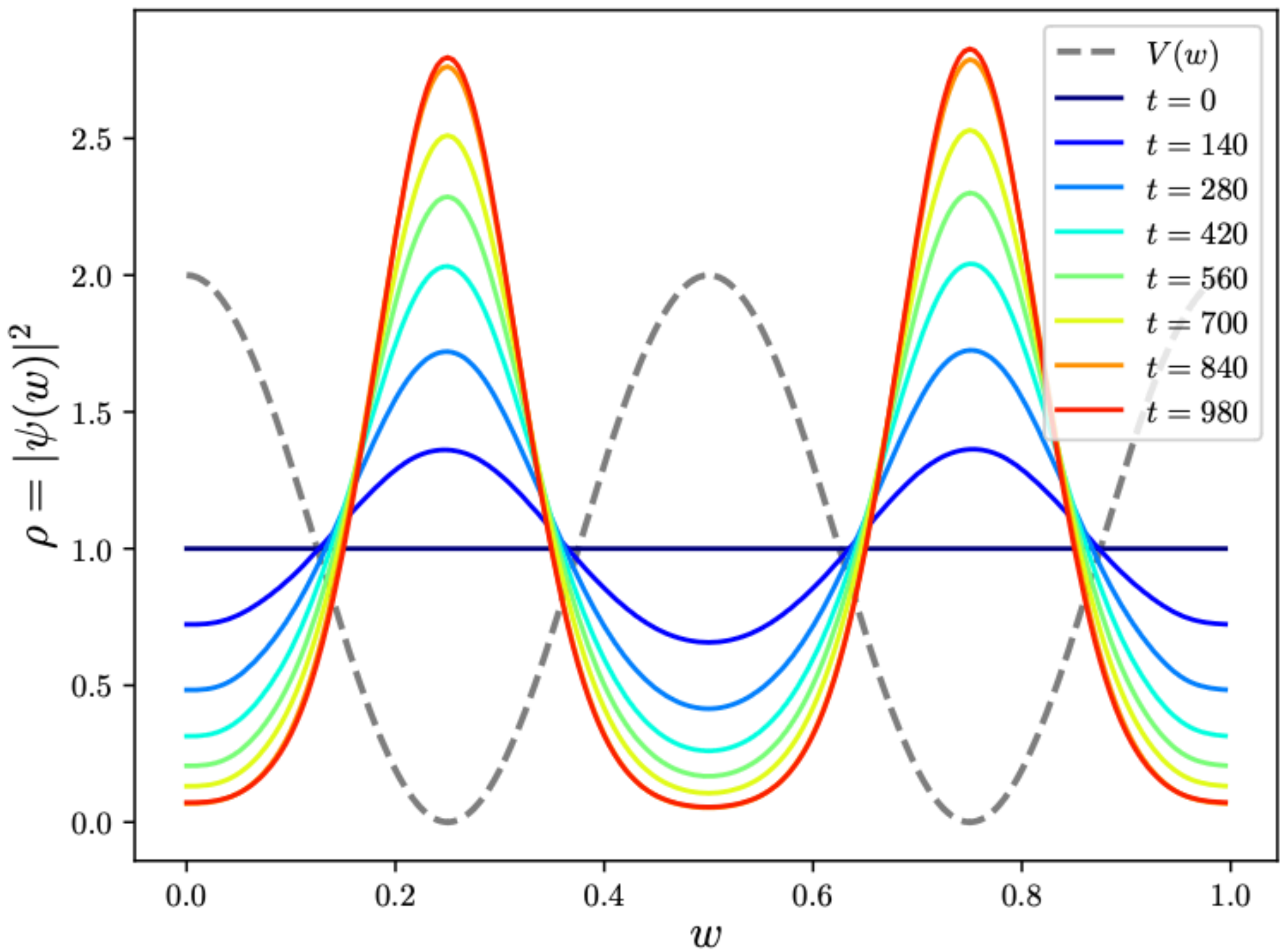
SAA, Criado, Spannowsky





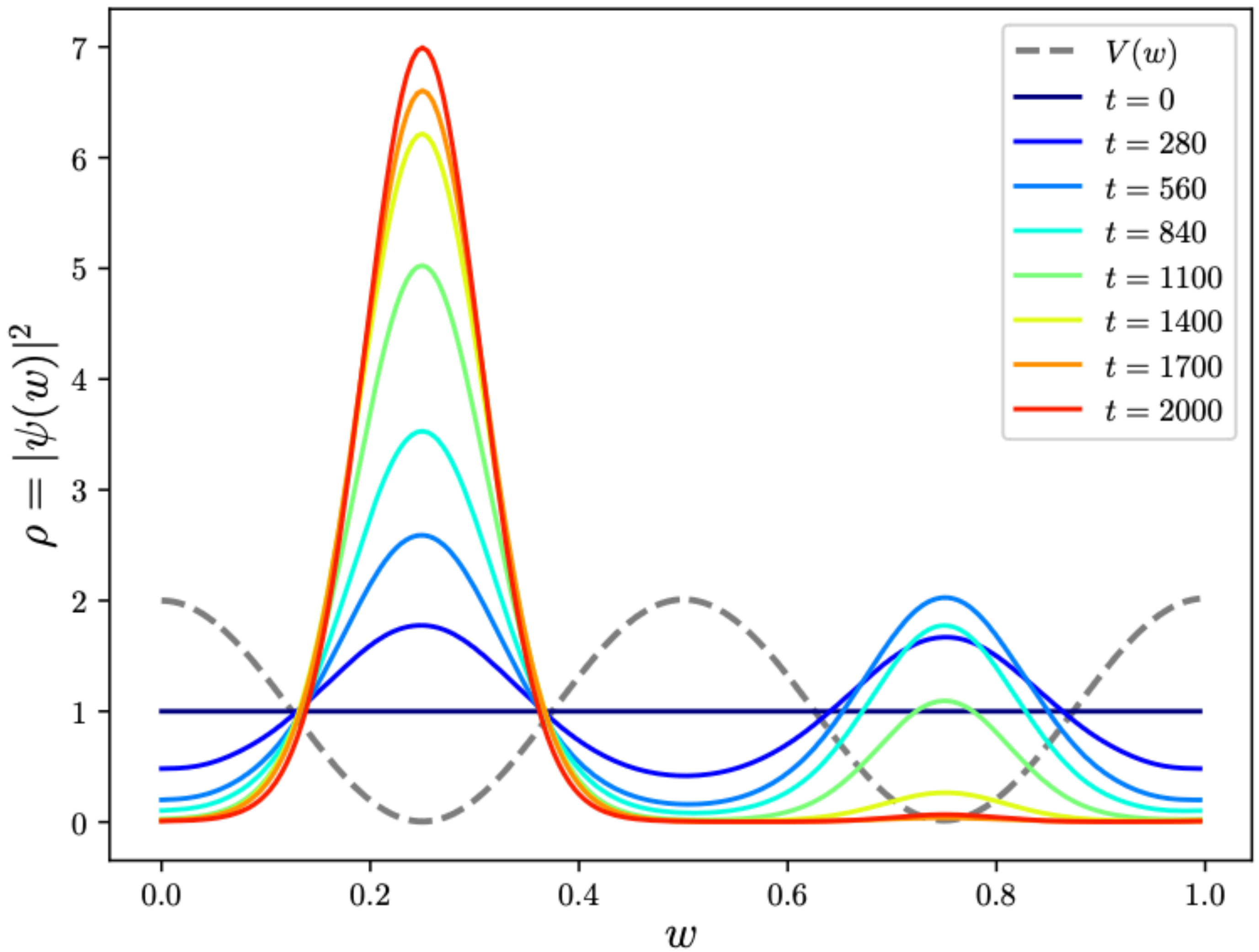
Tunnelling versus adiabatically evolving the ground state in a quartic potential. Here for tunnelling the initial wavefunction is chosen to be the groundstate of the approximate SHO potential around the false minimum. (Qubits binary-encode modes of truncated Hilbert space.)

What about a degenerate periodic cosine potential?



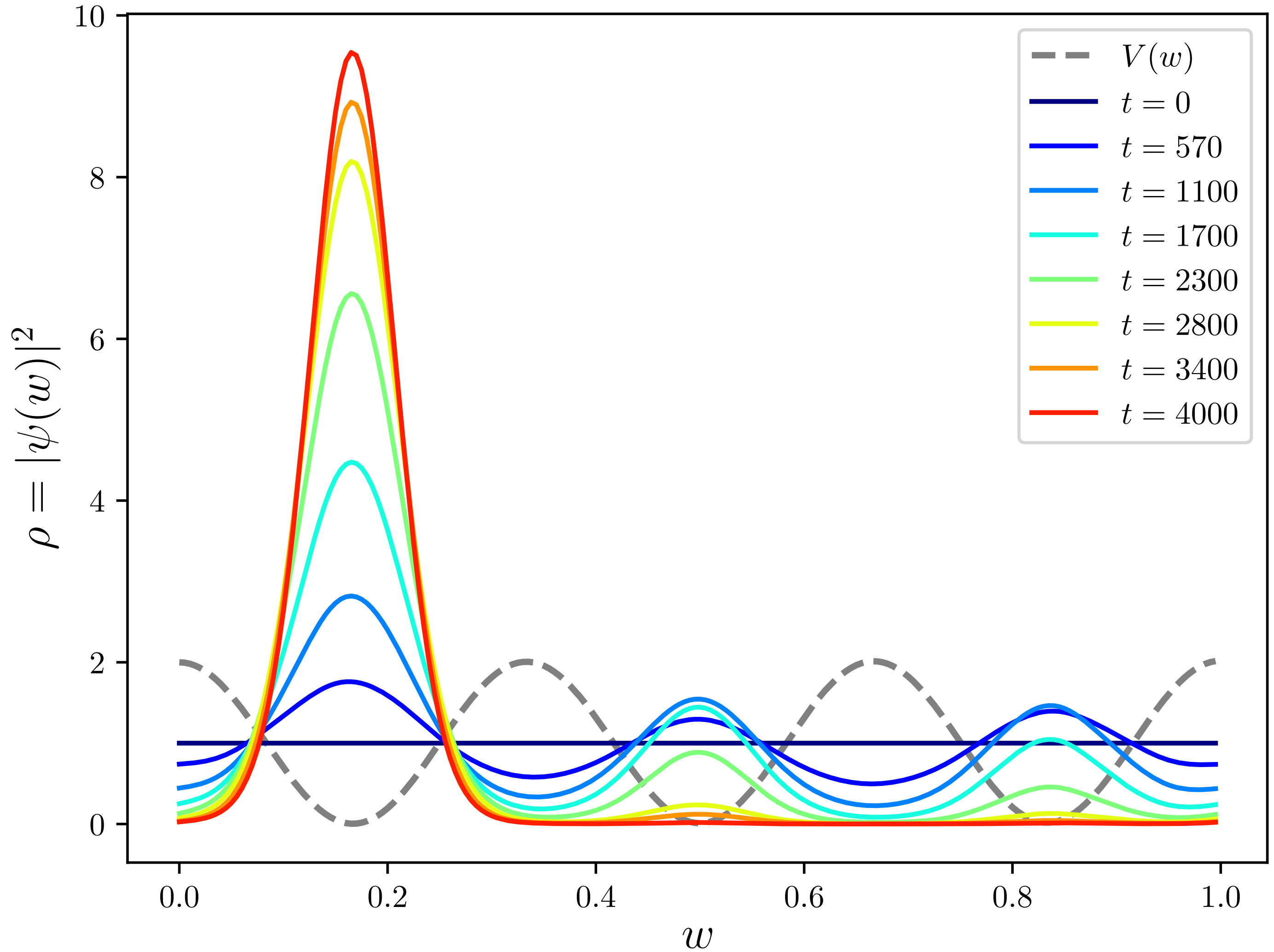
$$V(w) = 1 + \cos(4\pi w)$$

Very nearly but not quite degenerate cosine potential?



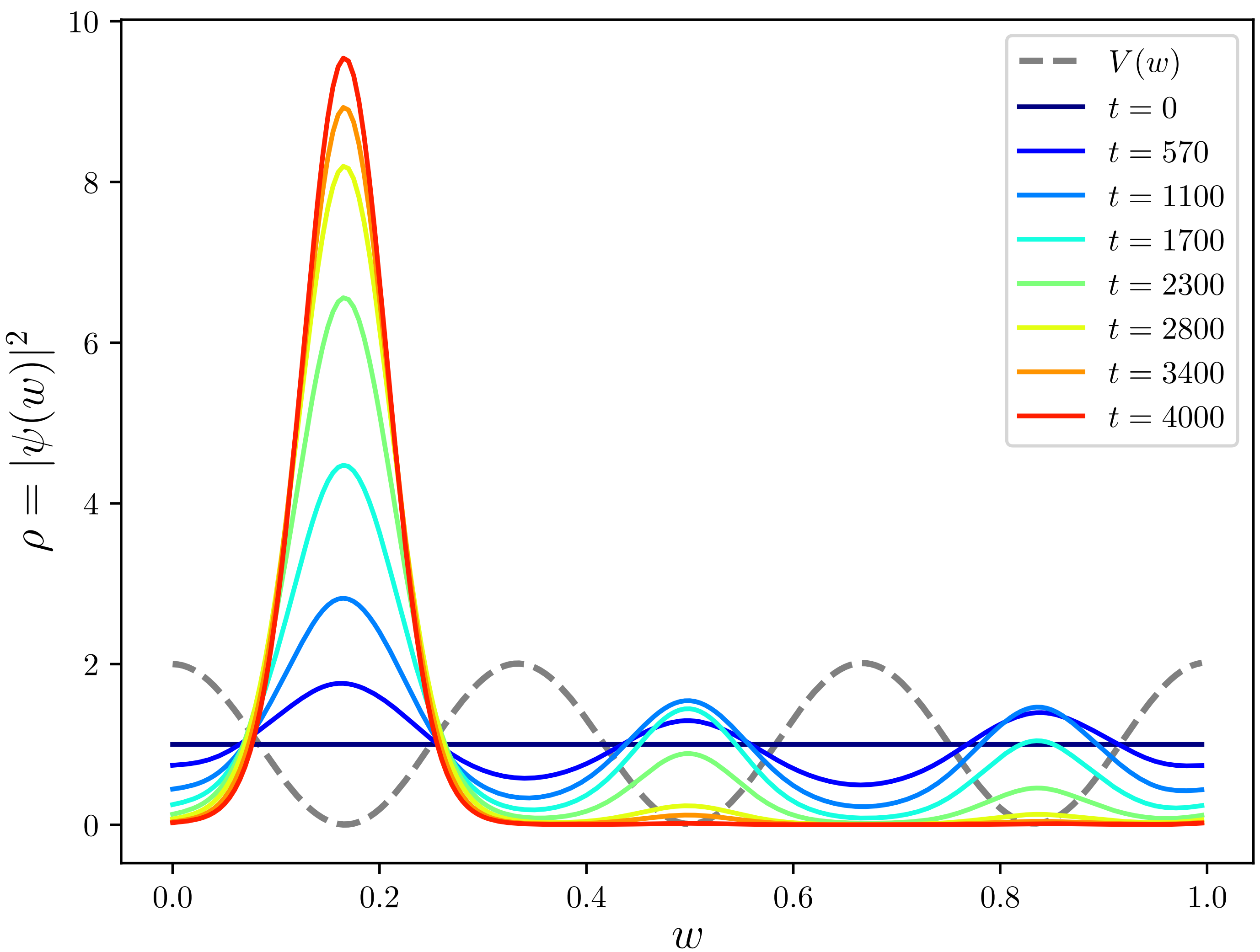
$$\Delta V_{\min} \approx 0.01$$

Very nearly but not quite degenerate periodic potential with 3 minima?



$$\Delta V_{\min} \approx 0.01$$

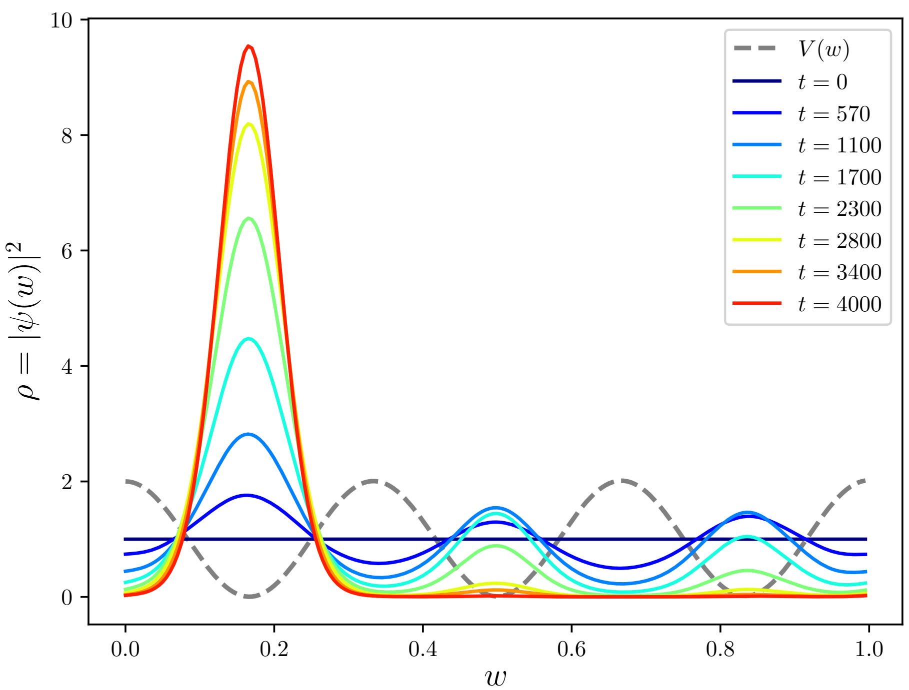
Very nearly but not quite degenerate periodic potential with 3 minima?



$$\Delta V_{\min} \approx 0.01$$

Maybe you can see where I am going with this …

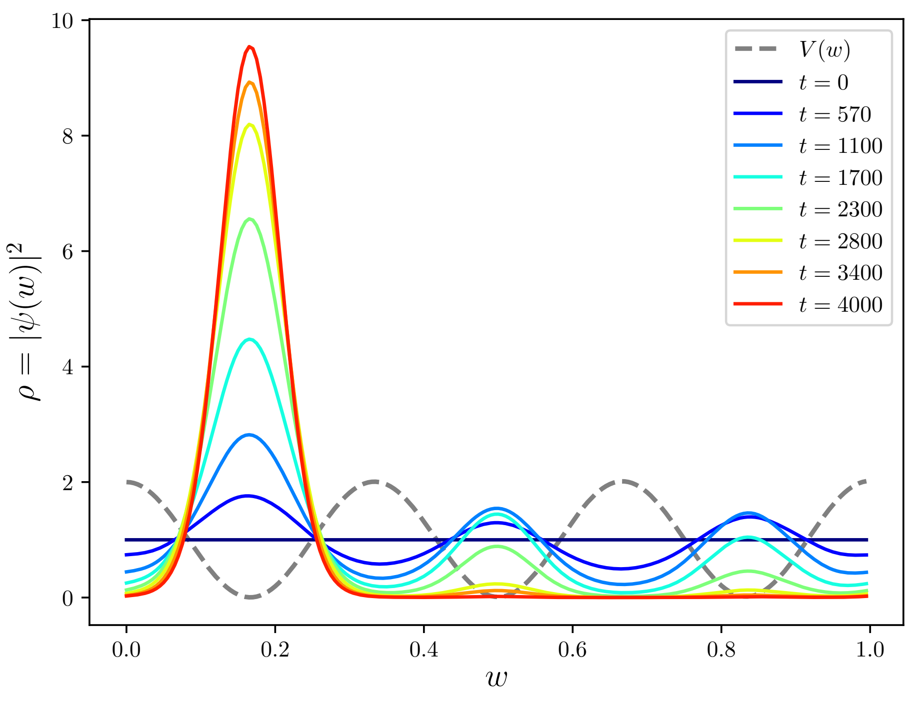
…back to the NMSSM domain wall problem.



is like selecting the NMSSM ``domain wall phase'': *If the potential terms that break to Z3 grow adiabatically, if we start in the groundstate even a tiny bias completely favours one vacuum and domain walls never form.*

Did we miss a get-out clause ?

$$\varepsilon \sim \lambda' \sigma M_W^2 / M_{\rm Pl}$$



Adiabaticity requires $\Delta H \times t \lesssim 1$

Take causal volume of size at least a domain wall width M_W^{-1}

Then $\Delta H = \varepsilon M_W$ gives

$$t \gtrsim \frac{M_{\rm Pl}}{\lambda' M_W^2} \approx \lambda'^{-1} 10^{-10} s$$

In conventional history, nucleosynthesis is at $t \approx 0.1 s$

So — answer from this back-of the envelope discussion is *not really* for the electroweak domain walls (as tradition dictates because I asked a question):

Conclusions about constraints on lambda' and which theories can satisfy the conditions without destabilising EW hierarchy (e.g. theories with R-symmetry) are similar with or without assumptions about the adiabatic evolution of the potential.

Steve: topological defects in an early period of adiabatic evolution of visible sector potentials seems an interesting possibility that is not often considered - in e.g. quintessence (c.f. Bean, Flanagan, Trodden; Denef and Douglas and the Bousso-Polchinski potential in string theory)

Subir: Nonsense. I bet there are at least 10 papers that discuss it. However none of them will work because they will all assume …

Conclusion: Happy 70th Subir …!

Thanks for your collaboration, support, being an oracle and your contributions to physics and life over the years!