Optimal control of a leaking qubit

P. Rebentrost\textsuperscript{1,2,3} and F.K. Wilhelm\textsuperscript{2}

\textsuperscript{1}Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, MA 02138, USA
\textsuperscript{2}IQC and Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada
\textsuperscript{3}Department Physik, Ludwig-Maximilians-Universit"{a}t, Theresienstr. 37, 80333 Munich, Germany

(Dated: August 20, 2008)

Physical implementations of quantum bits can contain coherent transitions to energetically close non-qubit states. In particular, for anharmonic oscillator systems such as the superconducting phase qubit and the transmon a two-level approximation is insufficient. We apply optimal control theory to the envelope of a resonant Rabi pulse in a qubit in the presence of a single, weakly off-resonant leakage level. The gate error of a spin flip operation reduces by orders of magnitude compared to simple pulse shapes. Near-perfect gates can be achieved for any pulse duration longer than an intrinsic limit given by the nonlinearity. The pulses can be understood as composite sequences that refocus the leakage transition. We also discuss ways to improve the pulse shapes.

PACS numbers: 02.30.Yy, 03.67.Lx, 85.25.Cp, 37.10.Jk

An ideal Hilbert space for a qubit has two dimensions, spanned by the states $|0\rangle$ and $|1\rangle$. However, in physical implementations quantum evolution can lead to transitions to non-qubit levels. In most cases the levels are energetically far away and hence are not the dominating limitation for quantum operations, for instance in the case of orbital degrees of freedom in nuclear or electronic spin-1/2 qubits \textsuperscript{1}. In other cases such as in superconducting qubits \textsuperscript{2} and optical lattices \textsuperscript{3} the situation is less favorable. In superconducting charge qubits the tunneling of more than one excess Cooper-pair on the superconducting island still has a sizable energy penalty \textsuperscript{4}. In phase qubits \textsuperscript{5,6,7}, optical lattices, and the transmon \textsuperscript{8,9} one has a third level which is only slightly detuned from the qubit energy splitting. These anharmonic oscillator qubits are the motivation for the present paper. Control strategies are required that go beyond the simplified two-level approximation. Initial phase qubit experiments \textsuperscript{5} used weak and thus slow Rabi pulses for qubit manipulations, making poor use of the available coherence time. Newer experiments \textsuperscript{7} combine Gaussian pulses with engineered spectrum and limited bandwidth. Initial proposals for avoiding leakage in superconducting qubits based on renormalization do not apply to phase qubits \textsuperscript{10} or lead to complex sequences of hard pulses \textsuperscript{11}.

Theoretical optimal control approaches are frequently used in quantum computing to synthesize quantum gates \textsuperscript{12,13,14}. The models mostly assume coherent evolution of coupled two-level systems such as nuclear spins \textsuperscript{15} or superconducting qubits \textsuperscript{16,17}. Progress also has been made in optimally controlling systems in the presence of decoherence mechanisms, either in Markovian \textsuperscript{18,19} or non-Markovian \textsuperscript{20,21} regimes, as well as an inhomogeneity \textsuperscript{22,23}. In this paper, we study a qubit in the presence of coherent transitions to a weakly off-resonant non-qubit level. We look for optimal pulses that generate quantum gates in the qubit subspace, assuming that a single control for the envelope of a resonant pulse is available. The pulse shapes significantly improve the fidelity obtained by refined standard techniques \textsuperscript{24}. We identify a time limit related to the nonlinearity above which efficient high fidelity pulses are possible.

Model — Our Hamiltonian of a qubit in the presence of a single leakage level is given by:

$$H(t) = \epsilon \sigma_z + \delta(t) \sigma_x + E_L |L\rangle \langle L| + \sqrt{2} \delta(t) (|1\rangle \langle 1| + h.c.). \quad (1)$$

Here, $\epsilon$ is the qubit level splitting while $\delta(t)$ is a time dependent control field. The Pauli matrices for the qubit are given by $\sigma_z = |1\rangle \langle 1| - |0\rangle \langle 0|$ and $\sigma_x = |1\rangle \langle 0| + |0\rangle \langle 1|$. Non-zero controls generate transitions not only in the qubit subspace but also to the leakage level $|L\rangle$ with energy $E_L$. Hence, conventional control techniques can reduce the quality of operations in the qubit subspace. This Hamiltonian approximates the three lowest energy states of a weakly nonlinear oscillator system, such as a Josephson phase qubit or a transmon \textsuperscript{24,25}. The matrix element of the transition $|1\rangle \rightarrow |L\rangle$ is scaled by $\sqrt{2}$. The detuning of the leakage level is the difference between the qubit splitting and the splitting of the transition $|1\rangle \rightarrow |L\rangle$, i.e. $\Delta \omega = 3 \epsilon - E_L$. Matrix elements between $|0\rangle$ and $|L\rangle$ are negligible following the parity selection rule of a nonlinear oscillator qubit at weak to intermediate detuning. For resonant driving on the qubit energy splitting, $\delta(t) = \lambda(t) \cos(\omega t)$ with $\omega = 2\epsilon$, one finds the Hamiltonian in the rotating wave approximation (RWA),

$$H^R(t) = -\Delta \omega |L\rangle \langle L| + \lambda(t) |1\rangle \langle 1| + \sqrt{2} \lambda(t) (|1\rangle \langle L| + h.c.). \quad (2)$$

A superscript $R$ denotes states and operators in the rotating frame, while $\lambda(t)$ is the envelope of the resonant pulse and can be viewed as a modulated Rabi frequency. The RWA representation assumes first that the logical qubits are encoded in the rotating frame; we drop the superscript $R$ for notational clarity. Second, it is assumed that the Rabi frequency $\lambda(t)$ is much smaller than $\epsilon$. This assumption is validated by results of this paper that show, that the Rabi frequency is ideally comparable with $\Delta \omega$. In the following, we will naturally focus on the qubit flip operation (NOT gate) given by,

$$U_F = e^{i\varphi_1} (e^{i\varphi_2} |L\rangle \langle L| + \sigma_x). \quad (3)$$
The global phase and the relative phase $\phi_1$ and the phase of the leakage level $\phi_2$ are meaningless for the NOT gate and thus taken to be arbitrary. First, we discuss controllability in the limits of weak ($\lambda \ll \Delta \omega$) and strong ($\lambda \gg \Delta \omega$) driving: for weak driving it is possible to make the population of the leakage level arbitrarily small by choosing the $\lambda$ small enough. The downside of this method is that the pulse has to be very long, therefore gate fidelity is limited by the coherence time. In the limit of strong driving, one in principle could make the pulses arbitrarily short and apply many of them within the coherence time. However, the strong driving Hamiltonian, in which $\Delta \omega$ is the smallest frequency, is similar to that of a single resonantly driven harmonic oscillator, i.e. $H(t) \approx \lambda(t)\sigma_x + \sqrt{2}\lambda(t)(|1\rangle\langle L| + |L\rangle\langle 1|)$. Hence, no high-amplitude control field creates a perfect NOT in the qubit subspace.

**Approximate solution** — One can find an approximate solution for our problem in the weak driving limit, $\lambda/\Delta \omega \ll 1$, that is exact to first order in $\lambda/\Delta \omega$. We denote by $W(t)$ a free evolution ($\lambda = 0$) of time $t$ and by $R(\theta)$ a weakly driven evolution, such that $\theta = \int_0^t \lambda(t')dt'$ with $\lambda/\Delta \omega \ll 1$. Then the solution is given by

$$U_F \simeq R(\theta_1) W \left( \frac{\pi}{2 \Delta \omega} \right) R(\theta_2) W \left( \frac{\pi}{\Delta \omega} \right) R(\theta_1).$$

This is a cascade of three pulses interrupted by two free evolutions of length $\pi/\Delta \omega$. Note that the free evolution for a time $t = \pi/\Delta \omega$ leads to an identity operation in the qubit subspace and a phase shift of $\pi$ on the leakage transition, i.e. $|0\rangle \rightarrow |0\rangle$ and $|L\rangle \rightarrow -|L\rangle$. Thus, in the isolated qubit subspace, we have just the three rotations with the condition $\theta_1 + \theta_2 = \pi/2$ (e.g. $\theta_1 = \pi/8$ and $\theta_2 = \pi/4$) for a NOT gate. The dynamics on the leakage transition can be identified with time-dependent perturbation theory after diagonalizing the qubit subspace in $[1]$ to lift the degeneracy. This pulse sequence removes transitions to the third level to first order in $\lambda/\Delta \omega$, with errors of the order $(\lambda/\Delta \omega)^2$ remaining. As a comparison, the rotation with only one weak pulse discussed in the previous paragraph introduces errors of the order $\lambda/\Delta \omega$. In principle, recursive application of such pulse sequences could remove the error up to any desired order, but would lead to a complicated and long pulse. Also, it needs to be pointed out that the rotations themselves require small $\lambda/\Delta \omega$ and thus inevitably take significant extra time beyond the waiting periods. In this work, we find optimal sequences with respect to fidelity and pulse duration. We resort to numerical methods of optimal control, namely the Gradient Ascent Pulse Engineering (GRAPE) algorithm [12].

**Control theory** — Given a Hamiltonian $H(t) = H(\lambda(t))$ such as Eq. (2), the goal is to find a function for the control parameter $\lambda(t)$, $t \in [0,t_g]$, such that a desired unitary quantum gate $U_F$ is generated. Formally, this can be achieved by minimizing the Euclidean distance between target and actual evolution, i.e. $\|U_F - U(t_g)\|_2^2 = \|U_F\|_2^2 + \|U(t_g)\|_2^2 - 2\text{Re} \text{tr} (U_F^* U(t_g))$. Here, $U(t)$ is the usual time evolution operator obeying the Schrödinger equation $\dot{U}(t) = -i(\lambda(t))U(t)$ with initial condition $U(0) = 1$. Minimizing the Euclidean distance is, up to a global phase, equivalent to maximizing the fidelity $\phi_1 = \frac{1}{2}\|\text{tr}(U_F^* U(t_g))\|^2$. The fidelity $\phi_1$ is not sensitive to the global phase $\varphi_1$, as required above. Additionally, we want to be insensitive to the relative phase of the leakage level. Averaging over the free phase $\varphi_2$ and keeping only the relevant terms leads to,

$$\phi_2 = \frac{1}{4} \left( \langle 0|U_F^* U(t_g)|0\rangle + \langle 1|U_F^* U(t_g)|1\rangle \right)^2.$$  (5)

In other words, it is sufficient to consider only the qubit subspace when evaluating the gate performance — everything else is fixed by unitarity of $U(t_g)$: maximizing $\phi_2$ automatically eliminates transitions to the third level. Optimization of fidelity $\phi_2$ requires the calculation of its gradient with respect to the controls. To this end, we introduce the approximate time evolution $U(t_g) \approx U_{1\cdots N} \ldots U_1$. The time interval $[0,t_g]$ is sliced into $N$ parts of length $\Delta t$, on each of which the controls and therefore the Hamiltonian is assumed to be constant. The propagator for an individual time step $U_j$ ($j = 1\ldots N$) can thus be written as $U_j = \exp \left( -\frac{i}{2} \Delta t H(j) \right)$, where $\lambda_j$ is the control amplitude during the $j$th time slice. Along the lines of [12], the gradient of the fidelity [5] with respect to $\lambda_j$ is given by,

$$\frac{\partial \phi_2}{\partial \lambda_j} = \frac{-i\Delta t}{2} \text{Re} \left\{ \sum_{k=0,1} \langle k|U_F^* U_{N\ldots j+1}|U_j \ldots U_1|k \rangle \sum_{m=0,1} \langle m|U_F^* U(t_g)|m \rangle \right\}.$$  (6)

This gradient is used in the GRAPE algorithm to find the maximum of the fidelity Eq. (5) as a function of the control parameters $\lambda_j$. Note that Eq. (6) is valid when the time slices $\Delta t$ are small compared to the characteristic time scales of the system.

**Pulses and performance** — The initial guess for the optimization is the $\pi/2$-pulse one would use in a two-level system, i.e. $\lambda = \pi/(2t_g)$ for all $t \leq t_g$. This pulse is then optimized using GRAPE based on the fidelity $\phi_2$. For now, we assume that the controls $\lambda(t)$ can take arbitrary values and that they can change arbitrary quickly from one time slice to the other. We assume further that we can permit arbitrary excursions to the limits of weak driving in the time $t_g$. As a result, at time $2\pi/\Delta \omega$ we recover a pulse similar to the approximate analytical solution, see Fig. (a). GRAPE finds a symmetric sequence of three soft yet high amplitude, approximately Gaussian pulses interrupted by two periods of free evolution. The areas under the three single pulse elements approximately correspond to the angles discussed earlier. Deviations are compensated by the negative amplitude during the free evolution parts. Fig. (b) shows the population of the qubit and the leakage level for the initial state being $|0\rangle$. The first pulse of the cascade transfers about 20% population to the excited state $|1\rangle$ while the leakage level remains almost unpopulated. After the first wait period, the second pulse populates the leakage level around 40%, where it remains during the second wait period. Here, the leakage
level accumulates a relative phase such that the third pulse leads to complete depopulation. For times below $t_{\text{opt}}$, pulses such as in Fig. 1 (c) are obtained. The pulse optimization tries to improve fidelities by rapidly turning the control field to very high amplitudes. However, the integral resource of the $2\pi/\Delta\omega$ waiting time is not available, which lowers the attainable fidelities. For longer gate times than $t_{\text{opt}} = 2\pi/\Delta\omega$ the optimal control field is a smooth pulse shape, see Fig. 1 (d). As an additional benefit, the control amplitudes are lower, leading to less transition to the non-qubit level during the application of the pulse.

Fig. 2 shows the gate error $1 - \phi_2$ versus pulse duration $t_g$ for unoptimized and optimized pulses. We see that GRAPE pulses easily achieve errors of $1 - \phi_2 = 10^{-6}$ and below. It has been estimated by various authors assuming different error models that the threshold error rate for fault-tolerant quantum computing lies between $10^{-2}$ and $10^{-6}$ [25, 27, 28]. Within the present model, pulse shaping easily reaches these thresholds for gate durations greater $t_{\text{opt}}$. Optimizing fidelity $\phi_2$ is advantageous over optimizing fidelity $\phi_1$: the irrelevant relative phase of the leakage level is not taken care of and therefore higher gate fidelities can be obtained. For comparison, we also show the gate error of standard pulse shapes like Gaussian and rectangular pulses [24]. Gaussian pulses used here are given by $\lambda(t) = \alpha/t_g \sqrt{\pi/2} \exp(-\alpha^2/(t_g^2(t - t_g/2)^2))$. These pulse shapes have attainable gate errors of $10^{-1} - 10^{-2}$ at the pulse durations considered here, orders of magnitudes worse than the optimized pulses. For rectangular pulses, only at gate times of $t_g > 280/\Delta\omega$ errors below $10^{-4}$ can be achieved. Essentially, these long gate times mean low control amplitudes and thus only small population of the leakage level: the weak-driving limit where $\lambda/\Delta\omega \ll 1$.

**Improving the pulse shape** — The pulses presented so far are obtained by solely optimizing the fidelity $\phi_2$, Eq. 5. As seen in Fig. 1 the controls have sharp initial rises, which may pose problems in the experimental implementation. In this section, we show that controls with smooth rises can be obtained at a small cost of pulse duration. We use the concept of penalty functions to constrain the gradient search algorithm. High amplitudes at the beginning and the end of the pulse can be penalized by amending the fidelity [5],

$$\tilde{\phi}_2 = \phi_2 - \int_0^{t_g} \gamma_A(t) \lambda^2(t) dt. \quad (7)$$

Here, we take the penalty strength as a function of time to be of the form

$$\gamma_A(t) = \gamma \left(2 - \tanh \left(\frac{t}{t_0}\right) + \tanh \left(\frac{t_g - t}{t_0}\right)\right), \quad (8)$$

where the positive $\gamma$ is the overall strength of the penalty and $t_0$ essentially parameterizes the rise time of the pulse. A non-zero $\gamma_A(t) \lambda^2(t)$ for any time $t$ will reduce the fidelity. A smooth penalty like in Eq. 8 leads to smooth pulses shapes. The gradient of $\tilde{\phi}_2$ is used in the optimization procedure.
Fig. 1 (e) shows the envelope of a pulse with $t_g = 10.0/\Delta \omega$, obtained by optimizing the fidelity (7) and (8) with the parameters $\gamma = 5.0/\Delta \omega$ and $t_0 = 0.1/\Delta \omega$. The control field starts at zero and increases within a time of around 0.5/\Delta \omega to the first peak of around $-0.1\Delta \omega$ and within a time of around 1.8/\Delta \omega to the second peak of around 0.6\Delta \omega. This pulse has a gate error below $10^{-4}$. In Fig. 2 the gate error of the amplitude-constrained pulses is demonstrated as a function of the pulse time. The parameters are again $\gamma = 5.0/\Delta \omega$ and $t_0 = 0.1/\Delta \omega$. The introduction of a penalty for the controls comes at a small cost of fidelity. The optimal gate time $t_{\text{opt}}$ of the unconstrained case is shifted to around 7.75/\Delta \omega. After that gate time, errors of below $10^{-6}$ are obtained, due to the fact that the $2\pi/\Delta \omega$ free evolution and the penalty requirements can be easily incorporated into a shaped control field.

**Discussion of the time scales** — We discuss the control fields and their properties in terms of the actual potential anharmonicity observed in the latest experiments in phase qubits and the transmon. In the phase qubit, the detuning is given in Fig. 5 of the supplementary material in [29] to be $\Delta \omega/2\pi = 0.2$ GHz. Thus, the optimal gate time is $t_{\text{opt}} = 5$ ns. This compares favorably to the 8ns FWHM Gaussian pulse employed in [29] with experimental gate fidelity of 0.98. The amplitudes of the pulse sequence at $t_{\text{opt}}$, Fig. 1(a), are around $\lambda/\Delta \omega = 0.8$ GHz, while the FWHM of the three approximately (half) Gaussian pulses is less than 1ns. Other pulses, e.g. Fig. 1(e) have a Rabi frequency and a modulation frequency of the order of the detuning. In the transmon, the detuning is slightly larger, $\Delta \omega/2\pi = 0.455$ GHz [9], leading to an optimal gate time of $t_{\text{opt}} = 2.2$ ns.

**Conclusion** — We have shown that high-fidelity quantum gates in a qubit can be performed despite the presence of a non-qubit leakage level. Our model approximates the resonantly driven Josephson phase qubit and the transmon, and is with modifications applicable to optical lattices. We have elucidated a composite sequence of pulses that performs a spin flip gate in the qubit subspace and refocusses the leakage transition to first order in the weak driving limit. Numerical optimization of the control field with the GRAPE algorithm drastically improves fidelities and pulse durations, leading to smooth, low-amplitude pulse shapes. We have identified an optimal pulse time $t_g \geq 2\pi/\Delta \omega$ that is integral to the problem and above which gate errors for a spin flip operation are below $1 - \phi_2 < 10^{-8}$. As a comparison, the rectangular pulse would take $t_g \approx 280/\Delta \omega$ for a gate error of $10^{-4}$. We have modified the pulse optimization with a penalty function such that the control fields become easier to realize in the experimental implementation. We have shown that this constraint can lead to similar near-perfect gate errors when the pulse duration is lengthened by about twice the required rise time.

We would like to acknowledge useful discussions with A.G. Fowler, J. Gambetta, J.M. Martinis, R. McDermott, F. Motzoi, and T. Schulte-Herbrüggen. This work was supported by NSERC through a discovery grant and QuantumWorks.