Parallel refreshed cryogenic charge-locking array with low power dissipation

X Bian[1],* G A D Briggs[1], and J A Mol[2]†
[1] *Department of Materials, University of Oxford, Oxford, OX1 3PH, UK*

[2] *School of Physics and Astronomy, Queen Mary University, London, E1 4NS, UK*

To build a large scale quantum circuit comprising millions of cryogenic qubits will require an efficient way to supply large numbers of classic control signals [1–5]. Given the limited number of direct connections allowed from room temperature, multiple level of signal multiplexing becomes essential. The stacking of hardware to accomplish this task is highly dependent on the lowest level implementation of control electronics [2], of which an open question is the feasibility of mK integration. Such integration is preferred for signal transmission and wire interconnection, provided it is not limited by the large power dissipation involved. Novel cryogenic electronics that prioritises power efficiency has to be developed to meet the tight thermal budget. In this paper, we present a power efficient approach to implement charge-locking array. As opposed to conventional approaches, where the power dissipation grows superlinearly with the total number of charge-locking units (quadratic growth with 1-dimensional addressing and to the power of $\frac{3}{2}$ with 2-dimensional cross-bar addressing as will be shown), our charge-locking scheme approaches linear power dissipation at large scale. The reduced power dissipation results from the parallel recharging method employed, which greatly decreases the number of switchings involved. To benchmark the power efficiency, we evaluate the power dissipation required to maintain $2^{14} \times 2^{14} \simeq 2.6 \times 10^8$ charge-locking units. As compared with serially refreshed charge-locking array with cross-bar addressing, our parallel refreshed charge-locking array shows more than 5000 times reduction in power dissipation and only dissipates 11 μW per kHz refreshing rate (assuming transistor gate size of 10 nm × 14 nm). Such a low power dissipation is compatible with the 1 mW cooling power available at 100 mK for large dilution fridges. We envision this highly efficient charge-locking scheme will lead to integrated classical control electronics for cryogenic quantum technologies.

PACS numbers:

I. INTRODUCTION

Semiconductor spin qubits based on gate defined quantum dot (QD) devices are promising candidates for building a universal quantum computer. They have the potential to reach high-level integration due to their small physical size and their compatibility with established semiconductor production processes. Although tremendous progress has been achieved over the last twenty years[6–16], a feasible path towards building a large scale quantum circuit based on QD devices remains to be established. The difficulty in scaling up lies not only in interconnecting (coherently coupling) between different QDs [9, 10, 17, 18], but also in interfacing with classic control signals (e.g. DC voltage, DC pulse, microwave pulse) that are required to operate (define, control and measure) QDs [1–5]. Classical control signal interfacing is conventionally supplied from room temperature electronics directly. Such a direct approach, albeit with high flexibility and simplicity, cannot be sustained as it depletes resources (e.g. cooling power, physical space) very quickly. A more efficient way to interface with classic control signals has to be implemented to facilitate upscaling.

A viable interfacing solution for scaling up quantum circuit requires fine tolerance levels for control signals and precise tuning to operate each QD. Large numbers of individually tunable signals have to be simultaneously maintained. Heat load and electrical noise have to be managed. Two different paths have been pursued in parallel towards developing a solution. One aims to minimise the variation of device characteristics, so that shared control lines can be used [19]. High-throughput electrical characterisation capability at cryogenic temperature can accelerate process optimisation [20–22]. Progress has been achieved in fabricating highly uniform double QDs[23], but the required level of uniformity remains to be achieved for a large array of QDs [24].

An alternative approach is to employ signal multiplexing circuitry and an array of charge-locking units [3, 25]. Each charge-locking unit is made up of a capacitor and a switch. The capacitor can be charged up to a static voltage when the switch is set to ON state and can hold that voltage for a period of time when the switch is set to OFF state. To maintain a voltage level, each capacitor has to be periodically recharged. A few proof-of-concept demonstrations [26–30] have been reported so far. The

*Electronic address: oums1005@materials.ox.ac.uk
†Electronic address: j.mol@qmul.ac.uk

schemes reported in [26] and [28] can be extended to charge-lock a large number of DC signals. The architecture suggested in [30] is also scalable by frequency multiplexing. In [26], signal multiplexing is accomplished by a cryogenic digital finite state machine (FSM) and digital clock to operate, which together contribute significant power dissipation, leaving sufficient cooling power for only ∼1000 charge-locked signals to be maintained. Further scaling will require some functionalities to be implemented at higher temperature to reduce the cryogenic heat load. An analog multiplexer (MUX) can control an array of charge-locking units using a multiple level selective gating (MLSG) method [28]. The number of multiplexed outputs depends exponentially on the number of control lines, and can rely on small number of control lines (reducing direct heat flow) supplied from control electronics implemented at higher temperature (mitigating heat load) to operate. The dependence of power dissipation on the total number of charge-locking units remains to be analysed.

In this paper, we will first analyse the power dissipation involved in operating this multiplexed 1-dimensional (1D) charge-locking circuitry. We will show that the power dissipated to perform periodic recharging is dominated by the MUX and grows quadratically with the total number of charge-locking units. Then the power dissipation analysis is extended to 2-dimensional (2D) charge-locking array implemented with conventional cross-bar addressing approach (see for example [3, 25]), which also shows superlinear dependence (to the power of $\frac{3}{2}$) on the total number of charge-locking units. Certainly, the accelerated growth rate of power dissipation with the size of charge-locking circuitry is not desired. Faced with this issue, we will propose a different scheme to charge-lock DC voltages at scale, where the signal lines for charging capacitors are multiplexed instead. A key improvement in our approach is that the power dissipation grows (very close to) linearly with the total number of charge-locking units (for both 1D and 2D addressing). The linear dependence results from the capability of our approach to perform periodic recharging in a parallel manner. Apart from the power dissipation aspect, improvement can also been achieved in the total time required to perform periodic recharging. This total recharging time is another critical aspect that can limit the size of the charge-lock circuitry. Lastly, we will generalise the MLSG method itself. As compared with base-2 multiplexing scheme, base-4 multiplexing scheme is as efficient in terms of scaling up the total number of multiplexed outputs, however will lead to further reduction in power dissipation.

II. POWER DISSIPATION ANALYSIS OF CONVENTIONAL APPROACH

Before presenting the detailed analysis on power dissipation, we will briefly explain the basic operation principle of the multiplexed charge-locking approach reported in [28]. It is worth explaining here this architecture is essentially the 1-dimensional equivalent of the cross-bar addressing suggested in [3, 25]. Figure 1a shows the schematics of a 2-level base-2 analog MUX based on the MLSG method. Physically, base-2 signal multiplexing corresponds to each channel (MESA) at current level being split into two channels at next level. The channel conduction is through the 2-dimensional electron gas (2DEG), which is normally on and can be suppressed by applying negative voltages to addressing gates. The selective channel gating is achieved by using polyimide to modulate gate capacitance. At each level, a pair of addressing gates are complementary, of which one controls the odd channels and the other controls the even channels. Therefore, each output can be individually selected by activating (applying negative voltage) a different combination of addressing gates. For example, output 0 is selected by activating addressing gate 1R/2R and output 1 is selected by activating addressing gate 1R/2L. To implement multiplexed charge-locking, each multiplexed output is connected to the gate of a transistor, which acts as the switch of a charge-locking unit. All the charge-locking units share a same input signal line V_{hold} for charging and the circuit schematics is shown in Figure 1b. The detailed process to initialise each charge-locking unit to hold a different voltage is described in [28] and will not be repeated here. We will focus on the process to perform periodic recharging. In the charge-locked state, all the multiplexed outputs are also charged-locked to hold a static voltage of V_g, which keeps the transistor of each charge-locking unit in OFF state and requires $2V_g$ to be applied to all the MUX addressing gates. V_g is determined by the most negative voltage to be charge-locked. To recharge a specific charge-locking unit (e.g. charge-locking unit 0), the corresponding multiplexed output is connected back to the MUX input (e.g. by deactivating 1L/2L, thus only 1R/2R are activated) and the input V_{in} is set to 0V. The transistor is thus turned on and the charge-locking unit can be recharged from the shared input V_{hold}. After recharging, the MUX input V_{in} is set back to V_g to turn off the transistor. All the deactivated addressing gates are then set back to $2V_g$ to put all the multiplexed outputs in charge-locked state. Same procedure can be repeated to recharge all other charge-locking units.

The power dissipated in periodically recharging the charge-locking circuitry has three components. First is the power consumed to recharge the capacitors that hold the static voltages (referred as holding capacitor C_H). Second is the power dissipated in driving the switches of the charge-lock units. The switch is implemented as a transistor, such that the power is essentially dissipated in charging/discharging the transistor gate (referred as transistor gate capacitor C_g). Third is the power dissipated in the signal multiplexing circuitry. Similar to the second contribution, the power consumed to oper-

ate the MLSG based MUX is essentially dissipated in charging/discharging the MUX addressing gates. The first contribution is constrained by the voltage resolution required (hence the minimum C_H) and the voltage drift δV_H allowed (i.e. the voltage swing involved in recharging). It has been shown previously in [27] that the second contribution can be orders of magnitude larger than the first contribution. However, the comparison is made with μm size transistor. We will strengthen this point here by showing that even if the transistor is implemented with the most advanced technology, the second contribution still dominates.

As the first two contributions are required for each single charge-locking unit, it is enough to compare them for one charge-locking unit. The first contribution P_H can be expressed as

$$P_H = C_H \delta V_H^2 f_c, \tag{1}$$

where f_c is the periodic recharging frequency. The voltage resolution is set by either the charge discreteness e/C_H or the thermal noise level $\sqrt{\frac{KT}{C_H}}$. To reach a voltage resolution of 1μV at a temperature of 100 mK, the holding capacitor C_H has to be larger than 1.4 pF, which is set by the thermal noise. With the conventional planar technology of 10 fF/μm^2, it might be 14 μm $\times$ 10 μm.

Similarly, the second contribution P_g can be expressed as

$$P_g = C_g \delta V_g^2 f_c, \tag{2}$$

where $\delta V_g = V_g - 0 = V_g$ is the voltage swing involved in turning on/off the transistor and C_g is the transistor gate capacitor. δV_g is largely determined by the voltage level required to operate QD and is on the order of $\sim$ 1V [3]. If the voltage drift δV_H to be compensated is 10 μV, the transistor gate has to be smaller than 0.14 nm $\times$ 0.1 nm to make P_g smaller than P_H, which certainly is not possible even with the most advanced technology. Suppose the transistor is implemented with a size of 14 nm $\times$ 10 nm, P_g is then at least 10^4 times larger as compared with P_H. As the transistor gate decreases, the threshold voltage will increase, thus the voltage V_g required to turn off the transistor will increase and the power dissipation will be even larger. This will be neglected in the following analysis, where we will approximate the total power consumption as the sum of the second and the third contributions.

For 1D charge-locking circuitry that consists of N charge-locking units being controlled by a K-level base-2 MUX (i.e. $N = 2^K$), the energy dissipated to sequentially switch on/off the transistor of each charge-locking unit once is

$$E_{S-SW1D} = NC_g(V_g - 0)^2 = NC_g V_g^2. \tag{3}$$

The capacitance of a MUX addressing gate is dominated by the area that is directly on top of the MESA, since the area on top of 500 μm thick substrate contributes negligibly and the area on top of the polyimide contributes at least 10 times smaller by design (as required by the selective gating method). In Figure 1a, the dominant contribution of gate capacitance is highlighted with red rectangles. The capacitance of each addressing gate is proportional to the number of channels that are effectively gated by it, and thus doubles for each higher level. To recharge a charge-locking unit involves one addressing gate at each level being switched. If one addressing gate is switched once at each level, the total energy dissipated in MUX to select one charge-locking unit is equal to

$$\begin{aligned} E_{S-MUX} &= (2^{K-1} + 2^{K-2} .... + 1) C_{MUX} (2V_g - 0)^2 \\ &= 4(N-1) C_{MUX} V_g^2 \end{aligned}, \tag{4}$$

where C_{MUX} is the capacitance contributed by each red rectangle. To sequentially recharge all the charge-locking units once, the total energy dissipated in analog MUX is thus equal to

$$E_{S-MUX1D} = NE_{S-MUX} = 4N(N-1)C_{MUX}V_g^2. \tag{5}$$

Suppose each red rectangle has the same capacitance as the transistor of each charge-locking unit, i.e. $C_{MUX} = C_g$. Then, for a recharging frequency of f_c, the total power dissipated is equal to

$$\begin{aligned} P_{S-1D} &= (E_{S-MUX1D} + E_{S-SW1D}) f_c \\ &= (4N^2 - 3N) C_g V_g^2 f_c \end{aligned}. \tag{6}$$

The power dissipated to periodically recharge the 1D charge-locking circuitry grows quadratically with the total number of charge-locking units. The major contribution comes from switching the highest level MUX addressing gates.

In a similar way, we can calculate the power dissipated to periodically recharge the 2D charge-locking array with cross-bar addressing (Figure 1c). Each charge-locking unit is connected to two transistors in series, of which one can be switched on/off by shared row control line and the other can be switched on/off by shared column control line. As a result, each combination of row/column control line corresponds one charge-locking unit. N column control lines and M row control lines are able to control $N \times M$ charge-locking units. As each column control line is shared by M transistors and each row control line is shared by N transistors, it can be obtained that the energy dissipated to sequentially switch on/off the two transistors of each charge-locking unit once is

$$\begin{aligned} E_{S-SW2D} &= NM(NC_g + MC_g)(V_g - 0)^2 \\ &= NM(N+M) C_g V_g^2 \end{aligned}. \tag{7}$$

For controlling large numbers of charge-locking units, row and column control lines need to be multiplexed.

The energy dissipated in MUX to select one row control line and one column control line once are E_{row} and E_{col} respectively, which can be expressed as

$$\begin{aligned} E_{S-row} &= 4(M-1)C_gV_g^2 \\ E_{S-col} &= 4(N-1)C_gV_g^2 \end{aligned} \tag{8}$$

following the reasoning used to obtain E_{S-MUX}. To sequentially recharge all the charge-locking units once, the total energy dissipated in two analog MUXs is thus equal to

$$\begin{aligned} E_{S-MUX2D} &= NM(E_{S-row} + E_{S-col}) \\ &= 4NM(N+M-2)C_gV_g^2. \end{aligned} \tag{9}$$

If the periodic recharging is performed at a frequency of f_c, then the power dissipation is

$$\begin{aligned} P_{S-2D} &= (E_{S-MUX2D} + E_{S-SW2D})f_c \\ &= 5NM(N+M-\frac{8}{5})C_gV_g^2f_c \end{aligned}. \tag{10}$$

For a charge-locking array of equal size in both dimensions i.e. $N = M$, the total number of charge-locking units is equal to N^2 and the total power dissipation P_{S-2D} has a superlinear dependence (to the power of $\frac{3}{2}$) on the total number of charge-locking units. As opposed to the 1D charge-locking circuitry, the power dissipated in MUX no longer dominates and is only roughly 4 times as much as that is dissipated in switching transistors of charge-locking units. In other words, the conventional cross-bar addressing itself will lead to a superlinear growth in power dissipation regardless of the exact implementations of signal multiplexing circuitry. The ratio 4 is based on the assumption that $2V_g$ being applied to MUX addressing gates is required to route a DC signal of V_g. It can be smaller in principle, though that will not affect the superlinear growth of power dissipation.

III. PARALLEL REFRESHED CHARGE-LOCKING APPROACH

In this section, we will present a different charge-locking approach, in which charge-locking units are recharged row-wise in a parallel manner, which greatly reduces the number of switchings required. The key difference between our approach and the conventional approach as described in Section II is that the multiplexed outputs are not used to drive the transistors (switches) of charge-locking units, but rather are used for charging the holding capacitors, with consequent power reduction. Unlike the conventional approaches, there is completely no difference in implementing 1-dimensional and 2-dimensional charge-locking unit addressing with our approach.

We will illustrate the basic charge-locking operation with a 2-level base-2 MUX, but the underlying principle can easily be extended to MUX of more levels. Figure 2a shows the column MUX used for performing parallel recharging, where each multiplexed output is connected to a recharging capacitor C_{Rj}, j = 0, 1, 2, 3. For 1D charge-locking, each multiplexed output is connected to a holding capacitor C_{H0j}, j = 0, 1, 2, 3, which can be simultaneously connected to column MUX for recharging and disconnected from column MUX for maintaining the static voltages, controlled by one shared control line e.g. GH_0 (Figure 2b). To extend to 2D charge-locking is essentially to add more rows of charge-locking units, where each row of charge-locking units share a control line to turn on/off transistors simultaneously (Figure 2c).

Figure 3 shows how each recharging capacitor of the column MUX is charge-locked to hold different static voltages, which is the core of our charge-locking approach (both initialising and recharging the holding capacitors). To initialise a specific row of charge-locking units to hold different voltages essentially follows the exactly same process except for connecting it to the column MUX beforehand, such that the same process will charge up both recharging capacitors and that row of holding capacitors (Figure 4a). It can then be disconnected from the column MUX (Figure 4b) without affecting the voltages being held while a different row of charge-locking units is being initialised. Charge-locking units can be initialised row by row to hold different static voltages (Figure 4c).

Key to the process shown in Figure 3 is to operate the addressing gates (i.e. built-in switches) of the column MUX in a sequence that keeps the previously charged capacitors disconnected. As result, the number of switchings is greatly reduced as compared with the conventional approaches described in Section II. The detailed procedure is as follows. We first activate addressing gate 1R/2R to select output 0 and set the input voltage to V_0, which will charge up capacitor C_{R0} (Figure 3a). We then activate addressing gate 1L and deactivate addressing gate 1R, which will disconnect capacitor C_{R0} from the input and select output 2. Subsequently, the input voltage is set to V_2 to charge up capacitor C_{R2}, while the static voltage at output 0 stays unchanged (Figure 3b). Thereafter, we activate addressing gate 2L and deactivate addressing gate 2R, to disconnect capacitor C_0 and C_2 from the input and select output 3. Up to here, output 0 and 1 are both in charge-locked state. Then the input is set to V_3 to charge capacitor C_{R3} (Figure 3c). Next we activate addressing gate 1R and deactivate addressing gate 1L, to disconnect capacitor C_{R3} from the input and select output 1. The input is set to V_1 to charge capacitor C_{R1} (Figure 3d). Lastly, we activate addressing gate 2R to isolate capacitor C_{R1} and C_{R3}. It is not essential to deactivate 1R here, the state in Figure 3e simply shows four capacitors each being charged to hold a different static voltage. This charge-locking approach can be easily extended to base-2 MUX of more levels. Table I shows the sequence to charge-lock 16 different static voltages with a 4-level base-2 MUX. Higher level gates, which have larger

capacitance, are switched less frequently. The approach demonstrated in [28] requires one addressing gate of each level to be switched once for any charge-locking unit to be charged. Our less frequent switching of higher level addressing gates reduces the power dissipation.

Next, we will describe the procedure to perform periodic recharging. The basic idea is illustrated in Figure 5a. Static voltages of V_R and $V_H - \delta V_H$ are held by recharging capacitor C_R and holding C_H respectively. When they are connected to each other, the charge redistributes among them, and an equilibrium voltage is reached and is equal to

$$V = \frac{V_R C_R + (V_H - \delta V_H) C_H}{C_R + C_H}. \tag{11}$$

If δV_H is the voltage to be compensated and to restore the original voltage requires $V = V_H$, then

$$V_R = V_H + \frac{C_H \delta V_H}{C_R}. \tag{12}$$

It might appear that the recharging capacitor C_R has to be at least as big as the holding capacitor C_H to precisely recharge the holding capacitor, but that is not necessary. To account for the effect of charge discreteness (Figure 5b), we introduce a new parameter N_H, which is the number of electrons leaked from the holding capacitor C_H and can be expressed as

$$N_H = \frac{C_H \delta V_H}{e}. \tag{13}$$

To ensure N_H flowing back to C_H during recharging, the recharging capacitor C_R should be charged up to

$$V_R = Round(V_H) + N_H \frac{e}{C_R}, \tag{14}$$

where $Round(V_H)$ is the voltage to be restored rounded to a coarse resolution $\frac{e}{C_R}$. For example, a voltage of 1411.1 μV rounded to a resolution of 10 μV will be 1410 μV. As discussed above, to reach a voltage resolution of 1 μV at 100 mK, the holding capacitor C_H has to be at least 1.4 pF and is set by the uncertainty introduced by thermal noise $\sqrt{\frac{KT}{C_H}}$. The resolution set by charge discreteness is $e/C_H \sim 0.1$ μV. If a recharging capacitor $C_R < 0.01 C_H$ is used for recharging, the voltage resolution of the isolated C_R is determined by charge discreteness and e/C_R will be larger than 10 μV. As will be shown later, the choice of C_R is relatively flexible for charge-locking circuitry with 1D addressing. For a 2D charge-locking array, a recharging capacitor much smaller than the holding capacitor reduces power dissipation.

Figure 5c shows each recharging capacitor C_{R0j} is charged to hold a static voltage V_{R0j} and prepared for recharging holding capacitors. Then all the transistors are turned on by the shared control line GH_0 and charge is redistributed to compensate for the voltage drift of each holding capacitor (Figure 5d). In practice, V_{R0j} here accounts for not only the voltage drift δV_{0j} arising from charge leakage (as determined by Equation 14) but also the systematic offset induced by charge injection from the transistor gate when the transistor is turned off, for each charge-locked voltage, which can be calibrated from the current at that specific charge-locked voltage [26–28]. The systematic offset induced by charge injection from the transistor gate depends on the relative size of the transistor gate and the holding capacitor, and for large transistor sizes can even be comparable to the voltage drift ([27]). If C_H and C_g are of size 10 μm $\times$ 14 μm and 10 nm $\times$ 14 nm respectively, C_H has a resolution of ~ 0.1 μV due to charge discreteness and C_g correspondingly has a resolution of ~ 100 mV. For $V_g \simeq 1$V applied to the transistor gate, about 10 electrons are confined in the transistor channel. If 5 electrons are being injected to C_H, 0.5 μV systematic offset is induced, which is smaller than the uncertainty induced by thermal noise.

Next, we will begin the analysis of the power dissipated in periodic recharging for both 1D and 2D charge-locking. For 1D charge-locking, suppose there are in total N charge-locking units, which are controlled by a K-level base-2 analog MUX (i.e. $N = 2^K$). As discussed in Section II, the total power dissipation still consists of two major contributions, it is easy to obtain that the energy dissipated to switch on/off transistors once is

$$E_{P-SW1D} = NC_g(V_g - 0)^2 = NC_g V_g^2. \tag{15}$$

To obtain the energy dissipated in operating the analog MUX, we observe that each i-th level addressing gate is switched on/off every 2^i outputs in the sequence to be set to charge-locked state (see Table I), of which the capacitance is $2^{i-1}C_g$. To prepare all the recharging capacitors, two addressing gates at each level will in total dissipate

$$\begin{aligned} E_{P-MUX1D} &= 2 \times K \times \frac{2^K}{2^i} \times 2^{i-1} \times C_g V_g^2 \\ &= KNC_g V_g^2 \end{aligned}. \tag{16}$$

The voltage applied to the addressing gates is V_g instead of $2V_g$, leading to further power reduction. The factor K accounts for the total K-level addressing gates. For a recharging frequency of f_c, the total power dissipated will be

$$\begin{aligned} P_{P-1D} &= (E_{P-SW1D} + E_{P-MUX1D}) f_c \\ &= (1+K) NC_g V_g^2 f_c. \end{aligned} \tag{17}$$

Equivalently, the total power dissipated can be expressed in the total number of charge-locking units N

$$P_{P-1D} = (1 + \log_2 N) NC_g V_g^2 f_c. \tag{18}$$

P_{P-1D} is very close to linearly growing with the total number of charge-locking units N, since $\log_2(N+1) - \log_2(N) \simeq 0$ for large N.

Output No	Binary	1L	1R	2L	2R	3L	3R	4L	4R
0	0000	OFF	ON	OFF	ON	OFF	ON	OFF	ON
8	1000	ON	OFF	OFF	ON	OFF	ON	OFF	ON
12	1100	ON	OFF	ON	OFF	OFF	ON	OFF	ON
4	0100	OFF	ON	ON	OFF	OFF	ON	OFF	ON
6	0110	OFF	ON	ON	OFF	ON	OFF	OFF	ON
14	1110	ON	OFF	ON	OFF	ON	OFF	OFF	ON
10	1010	ON	OFF	OFF	ON	ON	OFF	OFF	ON
2	0010	OFF	ON	OFF	ON	ON	OFF	OFF	ON
3	0011	OFF	ON	OFF	ON	ON	OFF	ON	OFF
11	1011	ON	OFF	OFF	ON	ON	OFF	ON	OFF
15	1111	ON	OFF	ON	OFF	ON	OFF	ON	OFF
7	0111	OFF	ON	ON	OFF	ON	OFF	ON	OFF
5	0101	OFF	ON	ON	OFF	OFF	ON	ON	OFF
13	1101	ON	OFF	ON	OFF	OFF	ON	ON	OFF
9	1001	ON	OFF	OFF	ON	OFF	ON	ON	OFF
1	0001	OFF	ON	OFF	ON	OFF	ON	ON	OFF

TABLE I: The sequence to charge-lock static voltages with a 4-level base-2 MUX. In the table, the activated state of addressing gate is labelled as ON and deactivated state of addressing gate is labelled as OFF. The relationship between the selected output and the state of addressing gates is given in binary representation. At each level, an activated L addressing gate corresponds to 1 and an activated R addressing gate corresponds 0. For example, output 11 is selected by activating 1L/2R/3L/4L, which is 1011 in binary and 11 in decimal.

Similarly, for a 2D charge-locking array that consists of M rows and N columns, i.e. of $N \times M$ charge-locking units in total. The energy dissipated in turning on/off the transistors of every charge-locking unit once is

$$E_{P-SW2D} = M \times E_{P-SW1D} = NMC_g V_g^2. \tag{19}$$

Then, the energy dissipated in column MUX is

$$\begin{aligned} E_{P-col} &= M \times E_{P-MUX1D} \\ &= NMKC_g V_g^2. \end{aligned} \tag{20}$$

As for the energy dissipated in row MUX, it can be obtained following the same reasoning to Equation 5

$$E_{P-row} = 4M(M-1)C_g V_g^2. \tag{21}$$

In addition to the above three contributions, there is a further contribution that is associated with preparing each recharging capacitor to the right voltage for recharging different holding capacitors, since the voltage held by each charge-locking unit varies along each column. If the RMS (root mean square) value of the voltage variation along each column is δV_R, then the total energy dissipated here is

$$E_{RC} = NMC_R \delta V_R^2. \tag{22}$$

Here, we have assumed the C_R is chosen to be sufficiently large, such that the voltage swing involved in preparing C_R for recharging different holding capacitors along the column is still dominated by the variation between each voltage being held.

For a recharging frequency of f_c and array of equal size in both dimensions, the total power dissipation is

$$\begin{aligned} P_{P-2D} &= (E_{P-SW2D} + E_{P-col} + E_{P-row} + E_{RC}) f_c \\ &= N^2(1 + \log_2 N + 4\frac{N-1}{N} + Q) C_g V_g^2 f_c \\ &\simeq N^2(5 + \log_2 N + Q) C_g V_g^2 f_c. \end{aligned} \tag{23}$$

where $Q = \frac{C_R \delta V_R^2}{C_g V_g^2}$ and can be minimised by choosing the smallest C_R possible. The choice of C_R is largely determined by δV_R. In an experiment on shuttling spin over a 9 QD array, the RMS value of the plunger gate voltage variation was calculated to be 47 mV (based on the Supplementary Table 1 in [31]). Below we will calculate a upper bound for Q using $\delta V_R = 100$ mV and a lower bound for Q using $\delta V_R = 10$ mV. Given the currently achieved variance of 47 mV and the demonstration of low disorder double QDs [23], $\delta V_R = 10$ mV is very likely to be achieved with further optimisation.

In the case of $\delta V_H = 100$ mV $\simeq \frac{1}{10} V_g$, the recharging capacitor C_R can be as small as $\frac{1}{1000} C_H$, such that a voltage drift of 10 μV in C_H corresponds roughly to 10 mV in C_R (see Equation 14), which is still an order of magnitude smaller than δV_H and ensures Equation 22 still being valid. Suppose C_g is of the size 14 nm $\times$ 10 nm, which is $\frac{1}{10^6} C_H = \frac{1}{10^3} C_R$, then Q is calculated to be 10.

Similarly, in the case of $\delta V_H = 10$ mV $\simeq \frac{1}{100} V_g$, the recharging capacitor C_R can be as small as $\frac{1}{100} C_H$, such that a voltage drift of 10 μV in C_H corresponds roughly to 1mV in C_R, which makes sure Equation 22 is still valid. Suppose C_g is of the size 14 nm $\times$ 10 nm, which

is $\frac{1}{10^6}C_H = \frac{1}{10^4}C_R$, then Q is calculated to be 1. From the value of Q, we conclude that this additional power dissipation involved in preparing recharging capacitor is comparable to other two contributions. P_{P-2D} is also very close to linear growth in power dissipation as the array size scales up. Unlike conventional approaches, improvements in uniformity lead to reductions in power dissipation.

IV. DISCUSSION AND CONCLUSION

Below we will evaluate the power dissipation required to maintain $2^{14}\times 2^{14} \simeq 2.6\times 10^8$ charge-locking units with both serially and parallel refreshed 2D charge-locking solutions, and benchmark them with respect to the cooling power of large dilution fridge available at 100mK, which is around 1 mW. It is generally accepted that order of 10^8 qubits are required for running error correction code to reach fault-tolerant quantum computation [3, 25]. The detailed calculation is shown in SI. For serially refreshed charge-locking solution with conventional 2D cross-bar addressing approach, the power dissipated to maintain 2.6×10^8 charge-locking units is around 61.5 mW per kHz refreshing rate. In contrast, the parallel refreshed charge-locking solution developed in this paper only dissipates 11 μW per kHz refreshing rate (assuming variation $\delta V_R \simeq 100$ mV and $Q \simeq 10$), which is more than 5000 times reduction in power dissipation. If the variation δV_R is improved to be around 10 mV, the power dissipation can be further reduced to be around 7.5 μW per kHz refreshing rate. As a comparison, an alternative solution to allow local control electronics residing closely with qubits is to operate them at 4K, of which the cooling power is around 1W and is thus only with 1000 fold increase. In other words, a parallel refreshed charge-locking array implemented at 100mK can be larger in size as compared with a serially refreshed charge-locking array implemented at 4K, solely from a power dissipation perspective.

Another more subtle advantage of our approach is related to the recharging frequency f_c. The minimum recharge frequency is determined by the leakage current and allowed static voltage drift. On the other hand, the time T required to recharge all the charge-locking units sets the upper limit for this recharge frequency, which in turn limits the total number of charge-locked signals that can be maintained. Similar reduction in recharging time is achieved with parallel refreshed charge-locking approach as compared with the conventional approaches (See SI for detailed analysis). For any charge-locking array with the size smaller than 10^{12}, it can be shown that the parallel refreshed charge-locking array show a constant $\frac{C_H}{C_R}$ fold reduction in total recharging time, essentially increasing the upper limit for charge-locking size by the same ratio.

Base-4 signal multiplexing based on MSLG method (Figure S1b) offers further reduction in the power dissipation compared with base-2 multiplex scheme. For the same number of control lines, a base-4 MUX scheme allows as many multiplexed outputs as base-2 signal multiplexing. The power reduction is obtained by replacing the $\log_2 N$ term in P_{P-2D} with a $\log_4 N$ term. As a result, it will lead to another 35 % reduction in power dissipation as compared with the base-2 addressing. This parallel refreshed charge-locking scheme can be generalised for any MLSG based MUX with arbitrary heterogeneous base stacking (See SI for two examples). With different base stacking, it adds another layer of flexibility to sub-divide the entire array, allowing different combinations of charge-locking units to be simultaneously selected. As a result, with additional local capacitors integrated, column-wise DC pulsing can be easily performed on specific rows as required by qubit operation.

To conclude, our parallel refreshed charge-locking approach yields significantly lower power dissipation than conventional approaches. Whereas conventional approaches all show superlinear dependence, in our system the power dissipation grows approximately linearly with the total number charge-locking units. The power dissipated to maintain $2^{14} \times 2^{14} = 2.6 \times 10^8$ charge-locking units is as low as 11 μW per kHz recharging frequency. The power dissipation can be further reduced with the continuing improvement in QD uniformity. For a voltage variation of $\delta V_H \simeq 10$ mV, it is expected to reach 7.5 μW per kHz recharging frequency. The total recharging time is also largely reduced with parallel recharging, which means one critical upper limit for the charge-locking array size is essentially increased. This charge-locking array requires minimal number of connections from higher temperature. In principle, 58 lines fed from higher temperature, 28 lines for the addressing gates of each MUX and 1 for the input of each MUX, will suffice to operate 100 million charge-locking units. In practice, there is another limit imposed by digital electronics (a few GHz switching rate). To resolve this problem, multiple independent inputs for Column MUX (as the structure used in [32]) can be used to reduce the required switching rate to a few GHz, which however will not change the overall power dissipation for the same array size. (See details in SI section A)

[1] C. G. Almudever, L. Lao, X. Fu, N. Khammassi, I. Ashraf, D. Iorga, S. Varsamopoulos, C. Eichler, A. Wallraff, L. Geck, et al., in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*

(IEEE, 2017).
[2] D. Franke, J. Clarke, L. Vandersypen, and M. Veldhorst, Microprocessors and Microsystems **67**, 1 (2019).
[3] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzurak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, and M. Veldhorst, npj Quantum Information **3** (2017).
[4] J. Hornibrook, J. Colless, I. C. Lamb, S. Pauka, H. Lu, A. Gossard, J. Watson, G. Gardner, S. Fallahi, M. Manfra, et al., Physical Review Applied **3** (2015).
[5] D. J. Reilly, npj Quantum Information **1** (2015).
[6] J. R. Petta, Science **309**, 2180 (2005).
[7] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, et al., Nature Nanotechnology **9**, 981 (2014).
[8] J. T. Muhonen, J. P. Dehollain, A. Laucht, F. E. Hudson, R. Kalra, T. Sekiguchi, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, et al., Nature Nanotechnology **9**, 986 (2014).
[9] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, et al., Nature **526**, 410 (2015).
[10] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, et al., Nature **555**, 633 (2018).
[11] D. M. Zajac, A. J. Sigillito, M. Russ, F. Borjans, J. M. Taylor, G. Burkard, and J. R. Petta, Science **359**, 439 (2017).
[12] W. Huang, C. H. Yang, K. W. Chan, T. Tanttu, B. Hensen, R. C. C. Leon, M. A. Fogarty, J. C. C. Hwang, F. E. Hudson, K. M. Itoh, et al., Nature **569**, 532 (2019).
[13] C. H. Yang, R. C. C. Leon, J. C. C. Hwang, A. Saraiva, T. Tanttu, W. Huang, J. C. Lemyre, K. W. Chan, K. Y. Tan, F. E. Hudson, et al., Nature **580**, 350 (2020).
[14] L. Petit, H. G. J. Eenink, M. Russ, W. I. L. Lawrie, N. W. Hendrickx, S. G. J. Philips, J. S. Clarke, L. M. K. Vandersypen, and M. Veldhorst, Nature **580**, 355 (2020).
[15] N. W. Hendrickx, D. P. Franke, A. Sammak, G. Scappucci, and M. Veldhorst, Nature **577**, 487 (2020).
[16] N. W. Hendrickx, W. I. L. Lawrie, M. Russ, F. van Riggelen, S. L. de Snoo, R. N. Schouten, A. Sammak, G. Scappucci, and M. Veldhorst, arXiv (2020), 2009.04268v1.
[17] X. Mi, J. V. Cady, D. M. Zajac, P. W. Deelman, and J. R. Petta, Science **355**, 156 (2017).
[18] N. Samkharadze, G. Zheng, N. Kalhor, D. Brousse, A. Sammak, U. C. Mendes, A. Blais, G. Scappucci, and L. M. K. Vandersypen, Science **359**, 1123 (2018).
[19] R. Li, L. Petit, D. P. Franke, J. P. Dehollain, J. Helsen, M. Steudtner, N. K. Thomas, Z. R. Yoscovits, K. J. Singh, S. Wehner, et al., Science Advances **4**, 3960 (2018).
[20] R. Pillarisetty, N. Kashani, P. Keys, R. Kotlyar, F. Luthi, D. Michalak, K. Millard, J. Roberts, J. Torres, O. Zietz, et al., in *2019 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2019).
[21] B. P. Wuetz, P. L. Bavdaz, L. A. Yeoh, R. Schouten, H. van der Does, M. Tiggelman, D. Sabbagh, A. Sammak, C. G. Almudever, F. Sebastiano, et al., npj Quantum Information **6** (2020).
[22] S. J. Pauka, K. Das, J. M. Hornibrook, G. C. Gardner, M. J. Manfra, M. C. Cassidy, and D. J. Reilly, arXiv (2019), 1908.07685v1.
[23] M. G. Borselli, K. Eng, R. S. Ross, T. M. Hazard, K. S. Holabird, B. Huang, A. A. Kiselev, P. W. Deelman, L. D. Warren, I. Milosavljevic, et al., Nanotechnology **26**, 375202 (2015).
[24] D. Zajac, T. Hazard, X. Mi, E. Nielsen, and J. Petta, Physical Review Applied **6** (2016).
[25] M. Veldhorst, H. G. J. Eenink, C. H. Yang, and A. S. Dzurak, Nature Communications **8** (2017).
[26] S. J. Pauka, K. Das, R. Kalra, A. Moini, Y. Yang, M. Trainer, A. Bousquet, C. Cantaloube, N. Dick, G. C. Gardner, et al., arXiv (2019), 1912.01299v1.
[27] Y. Xu, F. K. Unseld, A. Corna, A. M. J. Zwerver, A. Sammak, D. Brousse, N. Samkharadze, S. V. Amitonov, M. Veldhorst, G. Scappucci, et al., arXiv (2020), 2005.03851v1.
[28] R. K. Puddy, L. W. Smith, H. Al-Taie, C. H. Chong, I. Farrer, J. P. Griffiths, D. A. Ritchie, M. J. Kelly, M. Pepper, and C. G. Smith, Applied Physics Letters **107**, 143501 (2015).
[29] S. Schaal, S. Barraud, J. Morton, and M. Gonzalez-Zalba, Physical Review Applied **9** (2018).
[30] S. Schaal, A. Rossi, V. N. Ciriano-Tejel, T.-Y. Yang, S. Barraud, J. J. L. Morton, and M. F. Gonzalez-Zalba, Nature Electronics **2**, 236 (2019).
[31] A. R. Mills, D. M. Zajac, M. J. Gullans, F. J. Schupp, T. M. Hazard, and J. R. Petta, Nature Communications **10** (2019).
[32] D. R. Ward, D. E. Savage, M. G. Lagally, S. N. Coppersmith, and M. A. Eriksson, Applied Physics Letters **102**, 213107 (2013).

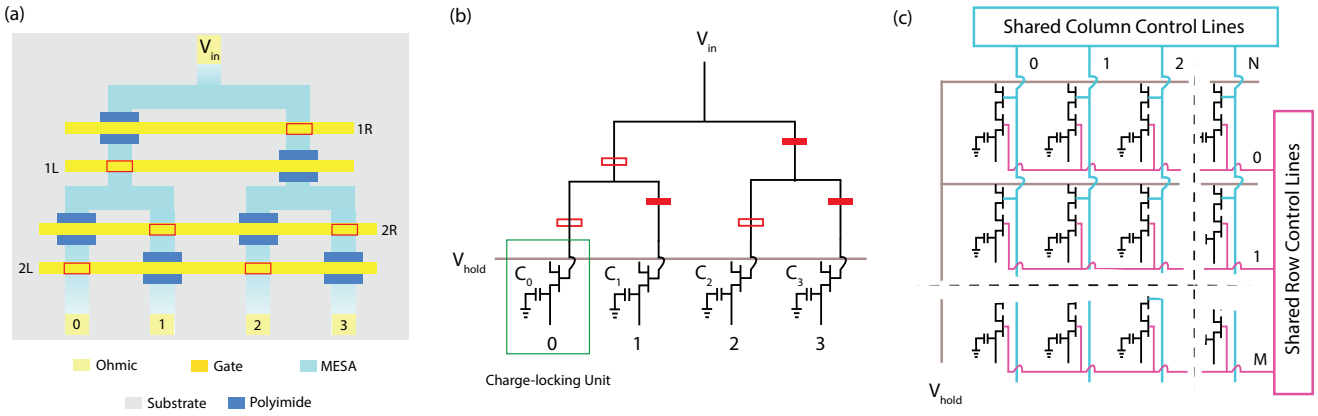

FIG. 1: (a) Schematics of a 2-level base-2 MUX, which relies on MLSG method to route DC signals. (b) Circuit schematics showing the MUX is used to drive the 4 transistors (i.e. switches) of charge-locking units, which is the approach taken in [28] and is the 1-dimensional equivalent of the cross-bar 2-dimensional addressing approach as proposed in [25]. All the charge-locking units share a same input V_{hold} for charging, such that each charge-locking unit can only be recharged in serially. In the circuit schematics, addressing gates are represented by horizontally aligned rectangles. Hollow red rectangles represent the addressing gates that are not activated and solid red rectangles represent the addressing gates that are activated. Charge-lock unit 0 is selected here. (c) Circuit schematics showing the 2-dimensional charge-locking array with the cross-bar addressing approach. Each charge-locking unit consists of a holding capacitor and a pair of transistors. The pair of transistors are controlled by shared column/row control line respectively, such that N column and M row control lines in total can address $N \times M$ charge-locking units. Each charge-locking unit can only be recharged sequentially.

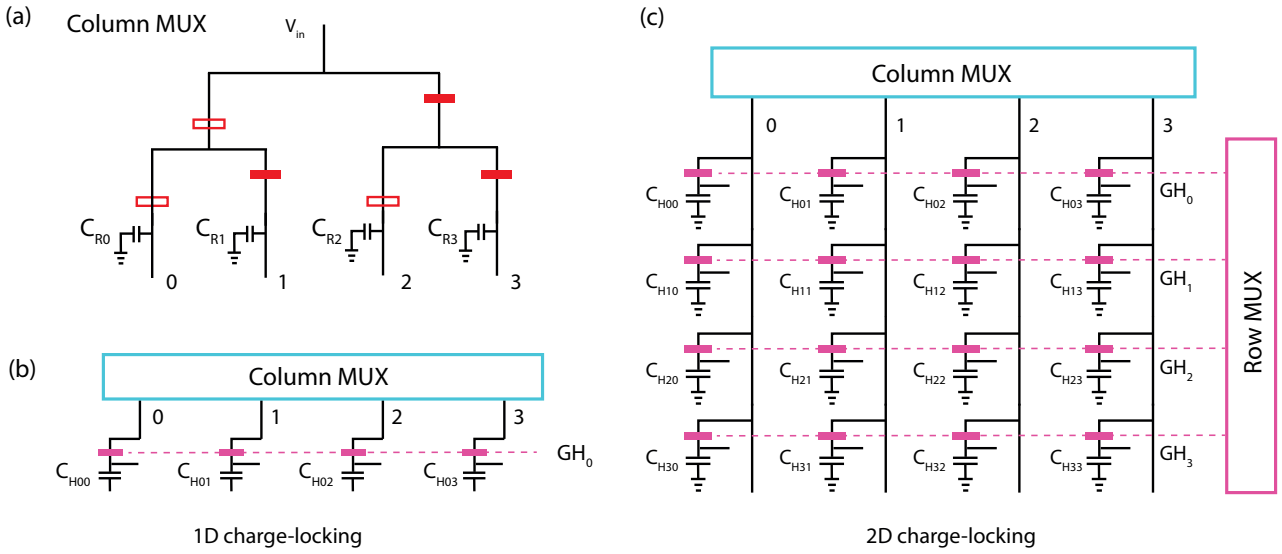

FIG. 2: (a) Column MUX (2-level base-2) for parallel recharging. The MUX itself is based on the MLSG method to route DC signals and each multiplexed output is connected to recharging capacitor C_{Rj} (j = 0, 1, 2, 3). (b) 1-dimensional charge-locking circuitry: As opposed to the conventional charge-locking approaches, where the switches of charge-locking units are multiplexed and the input line for charging is shared, the multiplexed outputs of Column MUX are each connected to a holding capacitor C_{H0j} (j =0, 1, 2, 3) and the switches are controlled by a shared control line. (c) To extend to two-dimensional charge-locking array, more rows of charge-locking units are added, where row-wise shared control lines are used for switching transistors. Row MUX is same as the Column MUX except that each multiplexed output is connected with a recharging capacitor.

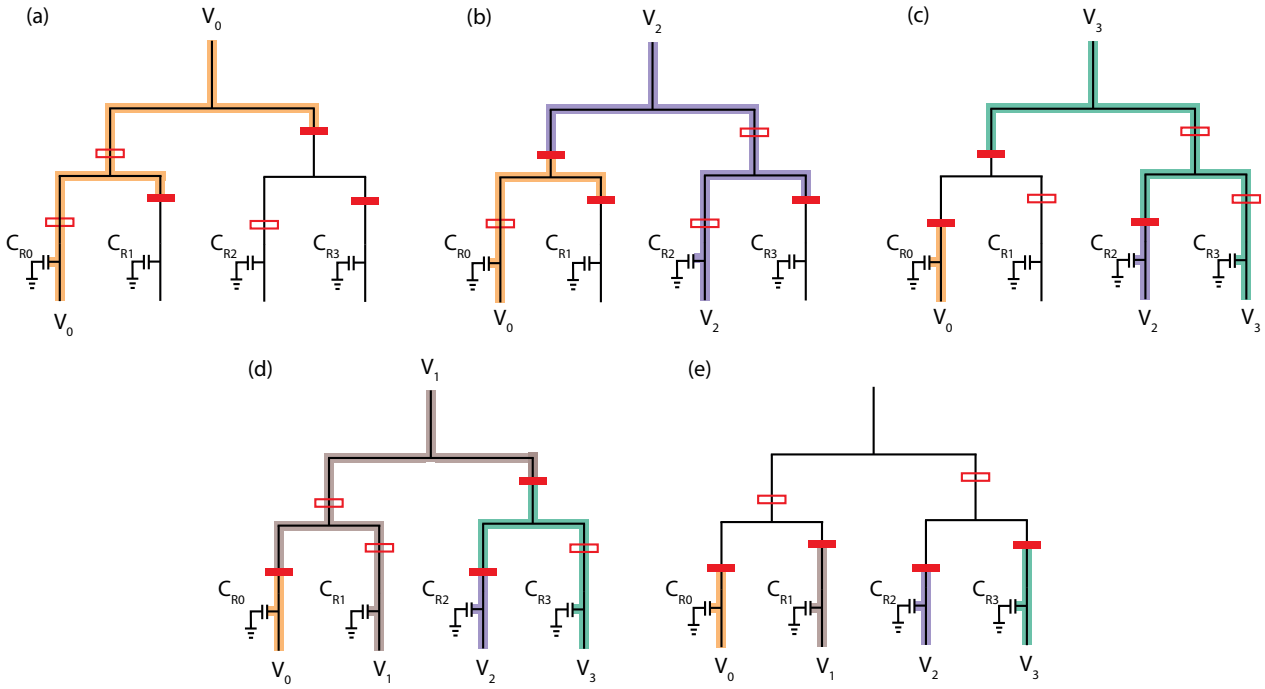

FIG. 3: Schematics showing the sequence to charge-lock 4 multiplexed output of Column MUX to hold different voltages, which is the core of our charge-locking approach and underpins both holding capacitor initialisation and parallel recharging process. By following this specific sequence of operating MUX addressing gates, the built-in switches of MUX allow simultaneously connecting one multiplexed output for charging while keeping the outputs that are already charged isolated. As a result, the total number of switchings is greatly reduced.

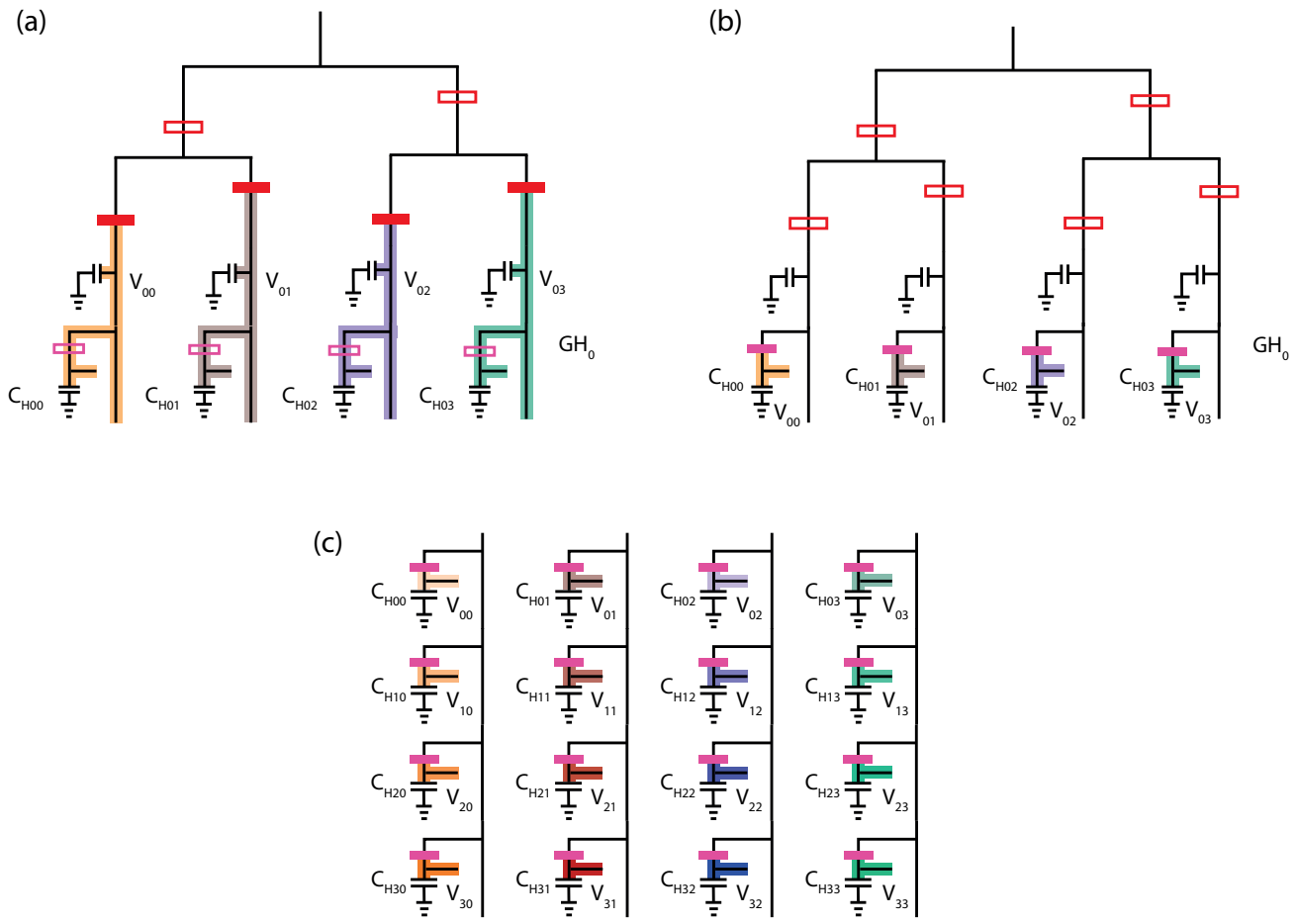

FIG. 4: (a) To initialise holding capacitor $C_{H0j}(j = 0, 1, 2, 3)$, transistors are turned by by the shared control line GH_0. Then each holding capacitor can be charged to hold a different voltage following exactly the same sequence as shown in Figure 3. (b) Transistors are turned off by the shared control line and all the holding capacitors are set into charge-locked state. (c) Each row of charge-locking units can be charged sequentially to hold different voltages by repeating steps shown in (a) and (b).

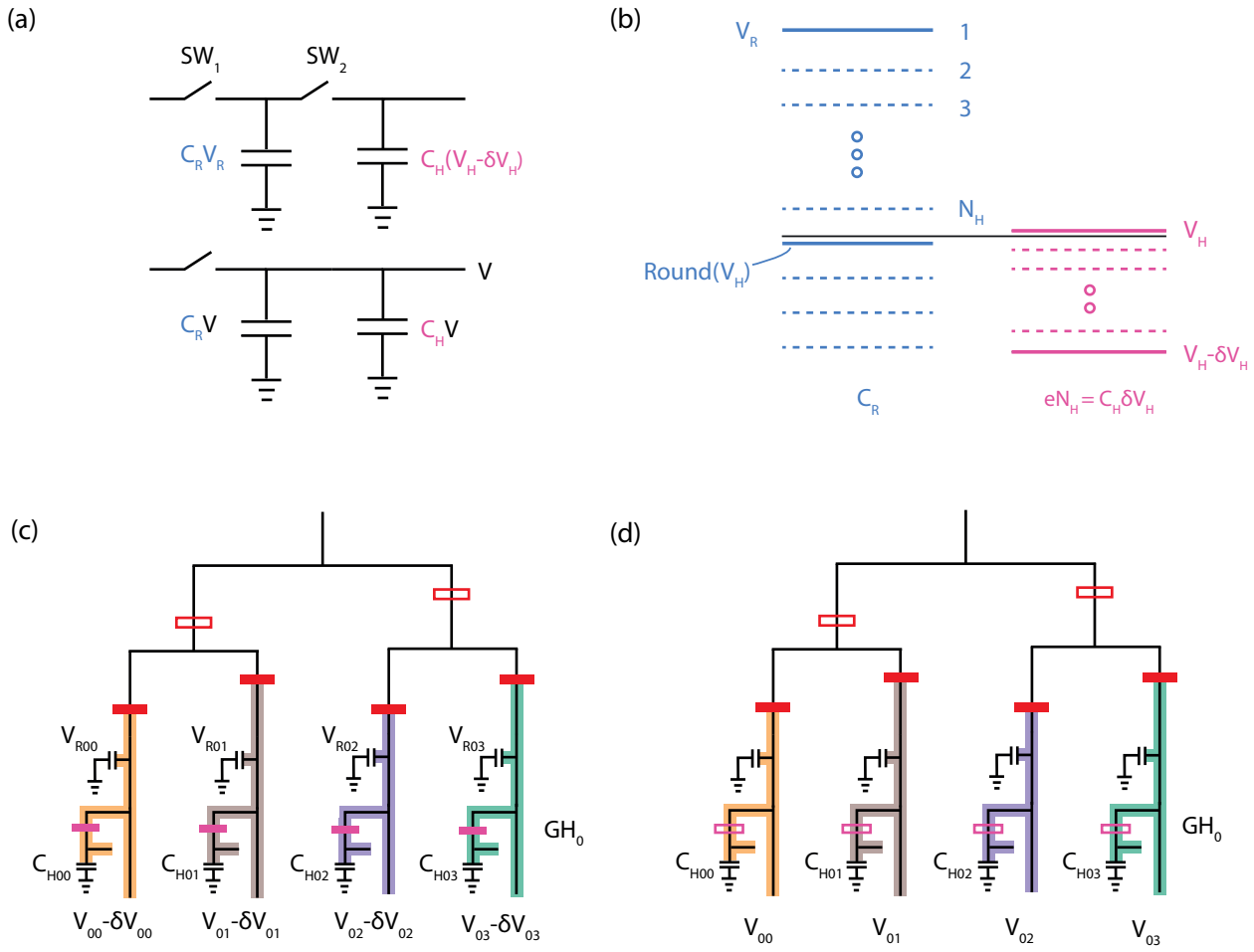

FIG. 5: (a) Recharging capacitor C_R and holding capacitor C_H are with two different voltages V_R and $V_H - \delta V_H$. Once they are connected with each other, charge redistributes among them, and an equilibrium voltage V is reached. This charge redistributing process can be employed to compensate the voltage drift δV_H in holding capacitor if C_R is set to the right voltage. (b) The voltage drift in C_H corresponds to a number of $N_H = \frac{C_H \delta V_H}{e}$ electrons to be compensated. To ensure N_H electrons flowing into C_H from C_R during charge redistribution, C_R should be prepared to a voltage level V_R, which is equal to $Round(V_H) + N_H e/C_R$. $Round(V_H)$ accounts for the charge-discreteness and is the highest voltage level of C_R below V_H. e/C_R is the voltage required to add each more electron to C_R. (c) All the holding capacitors $C_{H0_j} (j = 0, 1, 2, 3)$ are kept at charge-locked state, while the recharging capacitors are prepared to hold the right voltage for recharging. (d) The transistors are turned on by the shared control line GH_0 and charge redistributes among recharging capacitors and holding capactiors, such that holding capacitors are recharged in a parallel way.

V. SUPPLEMENTARY INFORMATION

A. $2^{14} \times 2^{14} \simeq 2.6 \times 10^8$ charge-locking units and transistor gate of 10 nm × 14 nm i.e. $C_g \simeq 1.4 \times 10^{-3} fF$

For serially refreshed charge-locking array with 2D cross-bar addressing, i.e. $N = M = 2^{14}$, the power dissipation is

$$\begin{aligned} P_{S-2D} &= 5NM(N + M - \frac{8}{5})C_g V_g^2 f_c \\ &\simeq 5 \times 2^{28} \times 2^{15} \times 1.4 \times 10^{-18} F \times 1V^2 \times 10^3 Hz \\ &= 61.5 mW/kHz \end{aligned} \tag{S1}$$

For parallel refreshed charge-locking approach with 2D addressing, i.e. $N = M = 2^{14}$ and $K = 14$, and assuming voltage variation $\delta V_R \simeq 100$ mV and $Q \simeq 10$,

$$\begin{aligned} P_{P-2D} &\simeq N^2(5 + \log_2 N + Q)C_g V_g^2 f_c \\ &\simeq 2^{28} \times (5 + 14 + 10) \times 1.4 \times 10^{-18} F \times 1V^2 \times 10^3 Hz \\ &= 11 \mu W/kHz \end{aligned} \tag{S2}$$

If the voltage variation δV_R is improved to be around 10mV, then $Q \simeq 1$, the power dissipation can be further reduced to

$$\begin{aligned} P_{P-2D} &\simeq N^2(5 + \log_2 N + Q)C_g V_g^2 f_c \\ &\simeq 2^{28} \times (5 + 14 + 1) \times 1.4 \times 10^{-18} F \times 1V^2 \times 10^3 Hz \\ &= 7.5 \mu W/kHz \end{aligned} \tag{S3}$$

Assume the refreshing rate is 1 kHz, the column MUX addressing gate and transistor (switch) of the charge locking unit will be switched at a rate of 100 GHz, which is not compatible with the digital electronics switching at a few GHz rate. To resolve this issue, multiple independent inputs can be used to column MUX. For example, column MUX can employ 16 independent inputs while reducing the MUX level from 14 to 10. As a result, the required switching rate is reduced from 100 GHz to 6.6 GHz met by digital electronics. The total required wire number will increase from 58 to 64. This problem is equally present in serially refreshed charge-locking array with cross-bar addressing. For the serially refreshed solution, this multiple independent inputs structure have to be employed for both column and row MUX.

B. Recharging time

Following the same process as power dissipation analysis, assuming the recharging time is limited by a simple RC time constant, where R is dominated by the output resistance of the control electronics at higher temperature.

For serially refreshed charge-locking approach with 2D cross-bar addressing,

$$T_{S-2D} \propto NMC_H + NMmax(N, M)C_g + NMmax(\frac{NCg}{2}, \frac{MCg}{2}) \tag{S4}$$

Here the first term accounts for the time related to recharge holding capacitors. The second term accounts for the time in switching the shared control line, since the column and row shared control are simultaneously switched, the longer time is used. The third term accounts for the time in routing the DC signals i.e. switching MUX addressing gates, similarly the longer one is taken between the Column MUX and the Row MUX. For an array with equal size in each dimension, i.e. $N = M$

$$\begin{aligned} T_{S-2D} &\propto N^2 C_H + N^3 C_g + \frac{N^3 C_g}{2} \\ &\propto N^2 C_H + \frac{3}{2} N^3 C_g \end{aligned} \tag{S5}$$

For parallel refreshed charge-locking array with 2D addressing, the time required to recharge a single row is

$$\begin{aligned} T_{P-1D} &\propto NC_g + NC_R + KNC_g \\ &\propto NC_g + NC_R + N\log_2 NC_g \end{aligned} \tag{S6}$$

Note, there is no term related to recharging holding capacitor here, since it is not limited by RC time constant and the time required for charge redistribution to occur is assumed be at much shorter time scale. The first term accounts for the time in switching transistors with the shared control line GH_0. The second term accounts for the time required to prepare the recharging capacitors. The third term accounts for the time in routing DC signals i.e. in switching MUX addressing gates.

Then the time required to recharge M rows in total is

$$T_{P-2D} \propto M \times T_{P-1D} + M \times \frac{MC_g}{2} \tag{S7}$$

The first term accounts for the time to recharge all rows of charge-locking units and the second term accounts for the time in routing DC signals by the row MUX, i.e. in switching addressing gates. For an array with equal size in each dimension, i.e. $N = M$

$$T_{P-2D} \propto N^2 C_R + N^2(\frac{3}{2} + \log_2 N)C_g \tag{S8}$$

Below we evaluate the recharging time, assuming realistic conditions. Suppose the transistor gate is of the size 10 nm $\times$ 14 nm, i.e. $C_g \simeq 10^{-6} C_H$.

For any charge-locking array with size smaller than 10^{12}, i.e. $N < 10^6$, the total recharging time for serially refreshed charge-locking array with 2D cross-bar addressing can be approximated as

$$\begin{aligned} T_{S-2D} &\propto N^2 C_H + \frac{3}{2} N^3 C_g \\ &\propto N^2 C_H \end{aligned} \tag{S9}$$

As the recharging capacitor C_R is also orders of magnitude larger than the transistor capacitor C_H in either situation, i.e. $C_R = 10^{-2} C_H$ or $C_R = 10^{-3} C_H$, the total recharging time for parallel refreshed charge-locking array can be approximated as

$$\begin{aligned} T_{P-2D} &\propto N^2 C_R + N^2(\frac{3}{2} + \log_2 N)C_g \\ &\propto N^2 C_R \end{aligned} \tag{S10}$$

C. Base-4 signal multiplexing

D. Generalised parallel refreshed charge-locking with MLSG based MUX

Below we show the specific sequence to operate MLSG based MUX as column MUX for parallel refreshed charge-locking, one for homogeneous base-3 and the other for heterogeneous base stacking.

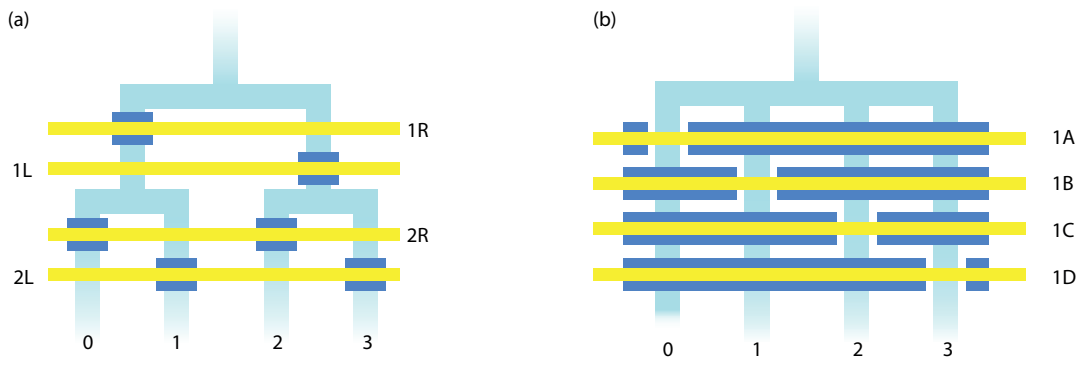

FIG. S1: (a) Base-2 singal multiplexing based on MSLG method, where each channel at current is split into two channels at next level. (b) Base-4 signal multipexing based on MSLG method, where each channel at current level is split into 4 channels at next level.

Output No	Bit Map	1A	1B	1C	2A	2B	2C	3A	3B	3C
0	000	OFF	ON	ON	OFF	ON	ON	OFF	ON	ON
9	100	ON	OFF	ON	OFF	ON	ON	OFF	ON	ON
18	200	ON	ON	OFF	OFF	ON	ON	OFF	ON	ON
21	210	ON	ON	OFF	ON	OFF	ON	OFF	ON	ON
12	110	ON	OFF	ON	ON	OFF	ON	OFF	ON	ON
3	010	OFF	ON	ON	ON	OFF	ON	OFF	ON	ON
6	020	OFF	ON	ON	ON	ON	OFF	OFF	ON	ON
15	120	ON	OFF	ON	ON	ON	OFF	OFF	ON	ON
24	220	ON	ON	OFF	ON	ON	OFF	OFF	ON	ON
25	221	ON	ON	OFF	ON	ON	OFF	ON	OFF	ON
16	121	ON	OFF	ON	ON	ON	OFF	ON	OFF	ON
7	021	OFF	ON	ON	ON	ON	OFF	ON	OFF	ON
4	011	OFF	ON	ON	ON	OFF	ON	ON	OFF	ON
13	111	ON	OFF	ON	ON	OFF	ON	ON	OFF	ON
22	211	ON	ON	OFF	ON	OFF	ON	ON	OFF	ON
19	201	ON	ON	OFF	OFF	ON	ON	ON	OFF	ON
10	101	ON	OFF	ON	OFF	ON	ON	ON	OFF	ON
1	001	OFF	ON	ON	OFF	ON	ON	ON	OFF	ON
2	002	OFF	ON	ON	OFF	ON	ON	ON	ON	OFF
11	102	ON	OFF	ON	OFF	ON	ON	ON	ON	OFF
20	202	ON	ON	OFF	OFF	ON	ON	ON	ON	OFF
23	212	ON	ON	OFF	ON	OFF	ON	ON	ON	OFF
14	112	ON	OFF	ON	ON	OFF	ON	ON	ON	OFF
5	012	OFF	ON	ON	ON	OFF	ON	ON	ON	OFF
8	022	OFF	ON	ON	ON	ON	OFF	ON	ON	OFF
17	122	ON	OFF	ON	ON	ON	OFF	ON	ON	OFF
26	222	ON	ON	OFF	ON	ON	OFF	ON	ON	OFF

TABLE S1: Sequence to operate a 3-level base-3 MUX for parallel refreshed charge-locking. Inactive addressing gates are labelled as OFF and active addressing gates are labelled as ON. For MUX of homogeneous base, there is a simple correspondence between the output channel and the inactive addressing gates. For example, 23 corresponds to 212 that is $23 = 2 \times 3^2 + 1 \times 3^1 + 2 \times 2^0$. Physically means all addressing gates other than gate C of level 1, gate B of level 2 and gate C of level 3 are activated to select output channel 23. Upon the transition from 18 to 21, 2nd level inactive addressing gate is changed from A to B. The activated addressing gate A of 2nd level keeps output channel 0, 9, 18 or equivalently 000, 100, 200 in charge-locked state. Upon the transition from 24 to 25, 3rd level inactive addressing is changed from A to B. The activated addressing gate A of 3rd level keeps output channel 0, 9, 18, 21, 12, 3, 6, 15, 24 or equivalently 000, 100, 200 210, 110, 010, 020, 120, 220 in charge-locked state.

Output No	Bit Map	1A	1B	2A	2B	2C	2D	2E	3A	3B	3C
0	000	OFF	ON	OFF	ON	ON	ON	ON	OFF	ON	ON
15	100	ON	OFF	OFF	ON	ON	ON	ON	OFF	ON	ON
18	110	ON	OFF	ON	OFF	ON	ON	ON	OFF	ON	ON
3	010	OFF	ON	ON	OFF	ON	ON	ON	OFF	ON	ON
6	020	OFF	ON	ON	ON	OFF	ON	ON	OFF	ON	ON
21	120	ON	OFF	ON	ON	OFF	ON	ON	OFF	ON	ON
24	130	ON	OFF	ON	ON	ON	OFF	ON	OFF	ON	ON
9	030	OFF	ON	ON	ON	ON	OFF	ON	OFF	ON	ON
12	040	OFF	ON	ON	ON	ON	ON	OFF	OFF	ON	ON
27	140	ON	OFF	ON	ON	ON	ON	OFF	OFF	ON	ON
28	141	ON	OFF	ON	ON	ON	ON	OFF	ON	OFF	ON
13	041	OFF	ON	ON	ON	ON	ON	OFF	ON	OFF	ON
10	031	OFF	ON	ON	ON	ON	OFF	ON	ON	OFF	ON
25	131	ON	OFF	ON	ON	ON	OFF	ON	ON	OFF	ON
22	121	ON	OFF	ON	ON	OFF	ON	ON	ON	OFF	ON
7	021	OFF	ON	ON	ON	OFF	ON	ON	ON	OFF	ON
4	011	OFF	ON	ON	OFF	ON	ON	ON	ON	OFF	ON
19	111	ON	OFF	ON	OFF	ON	ON	ON	ON	OFF	ON
16	101	ON	OFF	OFF	ON	ON	ON	ON	ON	OFF	ON
1	001	OFF	ON	OFF	ON	ON	ON	ON	ON	OFF	ON
2	002	OFF	ON	OFF	ON	ON	ON	ON	ON	ON	OFF
17	102	ON	OFF	OFF	ON	ON	ON	ON	ON	ON	OFF
20	112	ON	OFF	ON	OFF	ON	ON	ON	ON	ON	OFF
5	012	OFF	ON	ON	OFF	ON	ON	ON	ON	ON	OFF
8	022	OFF	ON	ON	ON	OFF	ON	ON	ON	ON	OFF
23	122	ON	OFF	ON	ON	OFF	ON	ON	ON	ON	OFF
26	132	ON	OFF	ON	ON	ON	OFF	ON	ON	ON	OFF
11	032	OFF	ON	ON	ON	ON	OFF	ON	ON	ON	OFF
14	042	OFF	ON	ON	ON	ON	ON	OFF	ON	ON	OFF
29	142	ON	OFF	ON	ON	ON	ON	OFF	ON	ON	OFF

TABLE S2: Sequence to operate a 3-level MUX of heterogeneous base (1st level of base2, 2nd level of base 5 and 3rd level of base 3) for parallel refreshed charge-locking. Inactive addressing gates are labelled as OFF and active addressing gates are labelled as ON. There is a one to one correspondence between the inactive addressing gates and the output channel. For example, output channel 26 is selected for charging by activating all addressing gates other than gate B of 1st level, gate D of 2nd level and gate C of 3rd level, that is $26 = 1 \times (3 \times 5) + 3 \times 3 + 1 \times 2$. Upon the transition from 15 to 18, 2nd level inactive addressing gate is changed from A to B. The activated addressing gate A of 2nd level keeps output channel 0, 15 or equivalently 000, 100 in charge-locked state. Upon the transition from 27 to 28, 3rd level inactive addressing is changed from A to B. The activated addressing gate 0 of 3rd level keeps output channel 0, 15, 18, 3, 6, 21, 24, 9, 12 or equivalently 000, 100, 110 010, 020, 120, 130, 030, 040, 140 in charge-locked state.