

A quantum annealer with fully programmable all-to-all coupling via Floquet engineering: Supplementary Information

This document is structured as follows:

- In Sup. Note 1, we describe our design principles, based on quantum Floquet theory, for programming arbitrary oscillator-oscillator couplings via dynamical control of a set of oscillators with a fixed set of all-to-all native couplings. We cast the problem as a generic, hardware-independent design problem of specifying phase-modulation indices for each oscillator. When the modulations required are small, we provide approximate solutions to the design problem in the form of analytic solutions at first order as well as an efficient numerical routine at second order. We also perform an analysis of the accuracy of the effective couplings and identify the dynamical-coupling parameter as a pivotal quantity in controlling the errors. Finally, we provide some guidelines for choosing the Floquet frequency in order to ensure that the Floquet approximation holds.
- In Sup. Note 2, we review the use of nonlinear Kerr parametric oscillators (KPOs) in quantum annealing to solve Ising spin problems. We present an archetypal model of a statically-coupled KPO network as introduced by Ref. [1], whose performance and solution mechanism we aim to approximate with our dynamical-coupling scheme. We also elaborate on some of the technical operational requirements of the statically-coupled annealer with respect to cavity-photon decay rate and problem-instance edge density, which we incorporate into our dynamically-coupled annealer.
- In Sup. Note 3, we present a hardware model for the realization of the KPO in the form of a Josephson parametric oscillator (JPO). We derive the relationship between the flux control signals applied to the JPO and resulting instantaneous frequency of the oscillator; the flux control signals are the control variables for achieving dynamical coupling. We also derive a model for an all-to-all-coupled network of JPOs using a common bus resonator, which provides the fixed native couplings that our dynamical-coupling scheme transforms into programmable effective couplings. We formulate and address the design problem that arises in performing dynamical coupling: what control signal should one apply to the flux lines in order to realize the dynamical modulations prescribed by our dynamical-coupling scheme given in Sup. Note 1?
- In Sup. Note 4, we analyze the structure of the native coupling matrix and how it is affected by both the dynamical modulations (which can lead to temporal variation in the couplings) as well as the bus-oscillator detunings of the oscillators (which affect the uniformity of the native couplings). Using this analysis, we give a prescription for choosing the bus-oscillator couplings and nominal detunings of each oscillator, in a way that ensures that the native coupling matrix possesses desirable properties such as approximate time-independence and uniformity while obeying all other constraints imposed by the analyses of previous sections. In particular, these properties dramatically simplify the procedure for solving the control design problem formulated in Sup. Note 3.
- In Sup. Note 5, we bring together the results of the above sections to give, for any desired Ising problem matrix, an explicit prescription for choosing the system parameters and modulations in order to realize dynamical coupling in a bus-mediated JPO network, while still respecting the constraints and approximations made in the above sections. With this prescription in hand, we then analyze the implications for scalability in the context of hardware limitations on the oscillator-bus detuning (i.e., system bandwidth) and the oscillator-bus coupling. We find that the main implication is that a maximum allowable nonlinear rate is imposed on the system, which translates into a minimum required cavity photon lifetime that generally increases with problem size.
- In Sup. Note 6, we describe the open-quantum system formalism we use to study the dynamics of the quantum annealer in our simulations. We also detail how we compute the metrics we use in our main manuscript figures, such as the success probability and spin configuration probabilities. We derive a classical model which is amenable to simulations of larger problem sizes in order to gain confidence that dynamical coupling can be achieved for more than a handful of oscillators.
- In Sup. Note 7, we analyze the decoherence and heating in the bus resonator. We present analytic expressions for the photon population in the bus resonator in the context of the dynamical-coupling scheme, and provide coherence-time requirements of the bus resonator as a function of problem size. In addition, we numerically study the effects of the bus decoherence on the performance of the quantum annealer by simulating a simplified model of the dynamically-coupled system that includes both the oscillator and bus modes.

- Finally, in Sup. Note 8, we provide additional data from our simulations which may be relevant or useful to interpreting our main results. In particular, we present the trajectories that resulted from simulating all 100 problem instances we used in our quantum simulations, as well as the correlation matrices for every problem size we considered in our classical simulations.

Finally, some notes on our conventions:

1. Throughout this work, we set $\hbar = 1$, so that energies (i.e., our Hamiltonians) are in units of frequency.
2. Furthermore, whenever the limits of a summation are not explicitly stated, it is implied that the index runs over the set $\{1, \dots, N\}$, where N is the problem size.
3. Finally, whenever we use the notation $x \sim y$, we mean the heuristic statement “ x is on the order of y ”.

Supplementary Note 1. Design principles for dynamical couplings

In the context of engineering dynamical couplings, we are primarily concerned with the problem of designing a set of control variables to cause a Hamiltonian possessing native couplings to approximate some other desired target Hamiltonian exhibiting static couplings of a particular form. More concretely, we suppose we start with a native Hamiltonian having the oscillatory form

$$\hat{H}_{\text{native}} = - \sum_{i \neq j} J_{ij}(t) \exp[-i(\phi_j(t) - \phi_i(t))] \hat{a}_i^\dagger \hat{a}_j, \quad (1)$$

where $J_{ij}(t)$ (satisfying $J_{ij} = J_{ji}$) is the magnitude of the native coupling between modes $\hat{a}_i$ and $\hat{a}_j$, and $\phi_i(t)$ are scalar functions of time. This Hamiltonian can be interpreted as the excitation-exchange energy between oscillators with nominal native couplings $J_{ij}(t)$ and where $\phi_i(t)$ is the total accrued phase of oscillator i in the Hamiltonian's frame. We show in Sup. Note 3 one realization of such a Hamiltonian in a system of nonlinear oscillators, in which time-dependent control of the functions $\phi_i(t)$ can be designed.

Our goal of engineering dynamical couplings is to achieve

$$\hat{H}_{\text{native}} \approx \hat{H}_{\text{target}} := -\lambda_C \sum_{i \neq j} C_{ij} \hat{a}_i^\dagger \hat{a}_j, \quad (2)$$

where C_{ij} (satisfying $C_{ij} = C_{ji}$ and $\max_{ij} |C_{ij}| = 1$) is the desired target coupling between modes $\hat{a}_i$ and $\hat{a}_j$, while the problem-strength parameter $\lambda_C \in \mathbb{R}^+$ controls the overall strength of the target couplings. In general, if we only have limited control over $J_{ij}(t)$ and $\phi_i(t)$, it is impossible to satisfy such a relationship (i.e., obtain $\hat{H}_{\text{native}} = \hat{H}_{\text{target}}$) for all times. To solve this problem, we utilize quantum Floquet theory [2]: if $\hat{H}_{\text{native}}$ is a periodic Hamiltonian with period $2\pi/\Lambda$, where $\Lambda \in \mathbb{R}^+$ is the Floquet frequency, then it can be approximated by its time average over one period:

$$\hat{H}_{\text{eff}} := \int_0^{2\pi/\Lambda} \hat{H}_{\text{native}}(t) \frac{dt}{2\pi/\Lambda}, \quad (3)$$

provided that Λ is sufficiently large (i.e., larger than any other energy scale in the system Hamiltonian) [2]. The evolution of the system under $\hat{H}_{\text{eff}}$ matches the evolution under $\hat{H}_{\text{native}}$ at periodic times $2\pi k/T$, where $k \in \mathbb{Z}$. For this reason, the dynamics under the ‘‘Floquet approximation’’ $\hat{H}_{\text{native}} \approx \hat{H}_{\text{eff}}$ is a stroboscopic approximation to the original dynamics. (Formally this is a leading-order approximation in the Magnus expansion of $\hat{H}_{\text{native}}$ [2].) We discuss in this section how we use the Floquet approximation to solve the design problem (2).

A. Floquet design problem

In principle, we can use any form of $\phi_i(t)$ that causes $\hat{H}_{\text{native}}$ to satisfy the Floquet approximation while achieving (2). For concreteness, however, we choose the following ansatz for the form of $\phi_i(t)$:

$$\phi_i(t) = \omega_0 t - (i-1)\Lambda t + \delta\phi_i(t), \quad \text{with} \quad (4a)$$

$$\delta\phi_i(t) := - \sum_{k=1}^{N-1} F_i^{(k)} \sin(k\Lambda t). \quad (4b)$$

This ansatz allows us to reduce the design problem for $\phi_i(t)$ into a design problem for the (scalar) Floquet design parameters $F_i^{(k)} \in \mathbb{R}$ for $i \in \{1, \dots, N\}$, $k \in \{1, \dots, N-1\}$. (In the main manuscript $F_i^{(k)}$ is alternatively named phase modulation coefficients.) Note that we also assume that $J_{ij}(t)$ can be made periodic with period $2\pi/\Lambda$. (Preferably, $J_{ij}(t)$ is approximately constant on the timescale of Λ .)

Interpretationally, if the energy shifts due to other terms of the Hamiltonian and $\delta\phi_i(t)$ are small, then the mean frequency of the i th oscillator is approximately $\omega_0 - (i-1)\Lambda$, corresponding to evenly-spaced detunings by the Floquet frequency Λ . The functions $\delta\phi_i(t)$ then represent additional dynamical phase modulation applied to each oscillator, which are used to engineer the effective interactions, with parametric controls represented by the scalars $F_j^{(k)}$.

Provided that the $\phi_i(t)$ have the form prescribed in (4), then the Floquet approximation tells us that

$$\hat{H}_{\text{native}} \approx \sum_{i \neq j} I_{ij} \hat{a}_i^\dagger \hat{a}_j, \quad (5a)$$

where the Floquet integral is

$$I_{ij} := \int_0^{2\pi/\Lambda} J_{ij}(t) \cos \left[(i-j)\Lambda t + \sum_{k=1}^{N-1} \left(F_i^{(k)} - F_j^{(k)} \right) \sin(k\Lambda t) \right] \frac{dt}{2\pi/\Lambda}. \quad (5b)$$

Thus, instead of trying to solve (2) directly for all times, we can instead try to satisfy

$$I_{ij} = \lambda_C C_{ij}. \quad (6)$$

This is the Floquet design equation, which results in (2) holding true in a time-averaged sense. We note that the time averaging does not preclude the possibility of designing a slow or adiabatic variation in, for example, λ_C (i.e., that it be a function $\lambda_C(t)$ of time), since any variation on timescales much longer than $2\pi/\Lambda$ is approximately constant in the Floquet integral; furthermore, such adiabatic variations need not be periodic in Λ .

The design parameters $F_i^{(k)}$ can be interpreted in a signal-processing sense as phase modulation indices since they cause the phase of the cosine function in (5b) to modulate in time. Thus, our scheme for achieving dynamical coupling can be equivalently viewed in a signal-processing framework as a phase-modulated coupling scheme.

B. Small-modulation approximation

One way to solve the Floquet design equation in general involves inserting the appropriate expression for J_{ij} and performing an optimization or root-finding search for $F_j^{(k)}$ by iterated evaluation of the integral. However, under certain special assumptions and/or restrictions, it is possible to simplify the problem further.

A number of assumptions can be made to simplify the Floquet equation. Here, we consider one such special case where 1) the time dependence of $J_{ij}(t)$ is negligible; and 2) the magnitude of the phase modulation terms satisfy

$$\varphi_{ij} := \sum_{k=1}^{N-1} \left(F_i^{(k)} - F_j^{(k)} \right) \sin(k\Lambda t) \sim \zeta, \quad (7)$$

where $\zeta \ll 1$. Then we can expand (5b) to first order in ζ (i.e., around $\varphi_{ij} = 0$) and (6) becomes, to first order,

$$-\frac{1}{2} J_{ij} \left(F_i^{(i-j)} - F_j^{(i-j)} \right) = \lambda_C C_{ij}, \quad (8)$$

for $i > j$. Thus, the problem reduces to a simple set of linear equations in $F_i^{(k)}$, which is readily solvable. We call this case the time-independent, small-modulation approximation. This equation likewise has an interpretation in terms of signal processing. In the rotating frame of the coupling interaction, the i th oscillator has carrier frequency $i\Lambda$, and so phase modulation at frequency $(i-j)\Lambda$ produces sidebands on oscillator i that overlap with oscillator j at carrier frequency $j\Lambda$.

There are a number of general solutions to (8), due to the symmetries of the equations. One solution prescribes that the i th oscillator be driven by $i-1$ different frequencies, according to the recursive relation

$$F_{k+j}^{(k)} = -\frac{2\lambda_C}{J_j^{(k)}} C_j^{(k)} + F_j^{(k)}, \quad (9)$$

with a base case of $F_i^{(k)} = 0$ for $1 \leq i \leq N$ and $i \leq k \leq N-1$. Here and elsewhere, $C_i^{(k)} := C_{i,i+k}$ and $J_i^{(k)} := J_{i,i+k}$ denotes the i th element of the k th diagonal of C and J , respectively.

Another solution, which emphasizes symmetry in the design parameters, is given by

$$F_i^{(k)} = \begin{cases} \sum_{m=0}^{\lfloor (N-k-i)/k \rfloor} \lambda_C C_{i+mk}^{(k)} / J_{i+mk}^{(k)}, & i \in \{1, \dots, \min(k, N-k)\} \\ -F_{[(i+k-1) \bmod k] + 1}^{(k)}, & i \in \{\max(N-k, k) + 1, \dots, N\} \\ F_{i-k}^{(k)} - 2\lambda_C C_{i-k}^{(k)} / J_{i-k}^{(k)}, & i \in \{k+1, \dots, N-k\} \\ 0, & \text{otherwise} \end{cases}, \quad (10)$$

As discussed later in Sup. Note 1D, this choice of solution centers the phase modulation strength on the middle oscillator ($i = N/2$), which makes the dynamical-coupling scheme more robust to noise. (See Sup. Note 1D for

further analysis of this solution.) Unless otherwise noted, this is the symmetry choice we will henceforth utilize for the first-order analytic solution.

While the first-order Taylor approximation provides an analytic form for the design parameters, we can further improve the approximation by expanding the Floquet integral to second order in ζ . At second order, (6) becomes

$$\begin{aligned} \frac{\lambda_C C_{ij}}{J_{ij}} = & -\frac{1}{2} \left(F_i^{(i-j)} - F_j^{(i-j)} \right) - \frac{1}{8} \sum_{l=1}^{N-1-(i-j)} \left(F_i^{(l)} - F_j^{(l)} \right) \left(F_i^{(i-j+l)} - F_j^{(i-j+l)} \right) \\ & - \frac{1}{8} \sum_{l=i-j+1}^{N-1} \left(F_i^{(l)} - F_j^{(l)} \right) \left(F_i^{(l-i+j)} - F_j^{(l-i+j)} \right) + \frac{1}{8} \sum_{l=1}^{i-j-1} \left(F_i^{(l)} - F_j^{(l)} \right) \left(F_i^{(i-j-l)} + F_j^{(i-j-l)} \right). \end{aligned} \quad (11)$$

In contrast to the first-order expansion (8), it is not possible in general to analytically solve this set of equations. However, as an algebraic system of second-order equations in $F_j^{(k)}$, it is still substantially easier to solve numerically than the full Floquet problem (6), and the residual errors are almost negligible even for realistic modulation depths (which stands in contrast to the situation for the first-order expansion). Further expansions in ζ are also possible, at the cost of algebraic complexity.

C. Numerical optimization for Floquet design parameters

As previously mentioned, the Floquet equation can be solved either by direct root finding or numerical optimization (specifically, minimization) over $F_i^{(k)}$ of the objective function

$$\sum_{i \neq j} \left(\lambda_C C_{ij} - \int_0^{2\pi} J_{ij} \cos \left[(i-j)t + \sum_{k=1}^{N-1} (F_i^{(k)} - F_j^{(k)}) \sin kt \right] \frac{dt}{2\pi} \right)^2, \quad (12)$$

given two matrices $\lambda_C C$ and J assumed to be time-independent here for simplicity, which allows us to perform a change of variables $t \mapsto t/\Lambda$ to scale out the Floquet frequency Λ . While direct root finding is generally prohibitively expensive, optimization of (12) usually works well for small N . However, this optimization occurs over $\mathcal{O}(N^2)$ variables (the number of $F_i^{(k)}$ entries), which degrades the quality of solutions for larger N (at least, for local methods and/or fixed optimization time).

To get around this issue, our preferred method for solving the Floquet equation is a “column-by-column” approach. This is inspired by the empirical observation that, given any particular set of design parameters $F_i^{(k)}$, modifying the values of the k th “column” $(F_1^{(k)}, F_2^{(k)}, \dots, F_N^{(k)})$ primarily (though not exclusively) produces changes in the k th *diagonal* of the Floquet integral $(I_{1,1+k}, I_{2,2+k}, \dots, I_{N-k,N})$ (and, symmetrically, in the $-k$ th diagonal). That is, instead of doing a single optimization over $\mathcal{O}(N^2)$ variables, we can opt to perform $\mathcal{O}(N)$ different optimizations, each over $\mathcal{O}(N)$ variables. Of course, because the various values of k are in fact coupled, this approach is unlikely to produce the global optimum. Nevertheless, we have found that error values of $|C_{ij} - I_{ij}/\lambda_C| \sim 10^{-4}$ are readily achievable for problem sizes $N \lesssim 100$, and it is reasonable to expect this behavior to continue for larger N as well.

To be more precise, while optimizing column k , we utilize the more restricted objective function

$$\sum_{j=1}^{N-k} \left(\frac{\lambda_C C_j^{(k)}}{J_j^{(k)}} - \int_0^{2\pi} \cos \left[\left(F_{j+k}^{(k)} - F_j^{(k)} \right) \sin kt + u_j^{(k)}(t) \right] \frac{dt}{2\pi} \right)^2, \quad (13a)$$

where the “cached” functions are

$$u_j^{(k)}(t) := kt + \sum_{l \neq k} \left(F_{j+k}^{(l)} - F_j^{(l)} \right) \sin lt, \quad (13b)$$

which contain all the values of $F_i^{(l)}$ that are not currently being optimized over and are thus static from the perspective of the k th-column optimization. (To see the connection between (13) and (12), consider $k := i - j$, so that $i = j + k$.) Thus, we need only update $u_j^{(k)}$ right before starting optimization over the k th column.

In this approach, we see that the objective function (13a) costs $\mathcal{O}(N^2)$ to evaluate: a factor of N for the summation and a factor of N for the integral, which needs $\mathcal{O}(N)$ samples in time. The updating of the cache functions $u_j^{(k)}(t)$

costs $\mathcal{O}(N^3)$: a factor of N for the summation, a factor of N for updating the value at every sample time t , and a final factor of N since there are $\mathcal{O}(N)$ different $u_j^{(k)}$ for each fixed k . Comparing (12) and (13), we improve the cost of the objective function by $\mathcal{O}(N)$, which may be significant depending on how the cost of the optimizer scales with the cost of the objective function, compared to the cost of the cache update.

As a technical point, we also note that in (13a), there are only $N - k$ elements in the summation, but as written, there are N different values of the Floquet design parameters $F_i^{(k)}$; this redundancy can be eliminated by a choice for the symmetry of the design parameters. For example, in the form chosen for (10), only $(F_1^{(k)}, \dots, F_{N-k}^{(k)})$ are unique, while all other values in that column are either zero or a function of those $N - k$ values.

Similarly, this column-by-column approach can also be adapted to perform numerical optimization in solving the second-order equations (11). In this case, the objective function is

$$\sum_{j=1}^{N-k} \left(\frac{\lambda_C C_j^{(k)}}{J_j^{(k)}} + \frac{1}{4} (F_{j+k}^{(k)} - F_j^{(k)}) \left[2 + \theta_{2k}^{N-1} (F_{j+k}^{(2k)} - F_j^{(2k)}) \right] + u_j^{(k)} \right)^2, \quad (14a)$$

where the binary-valued step function $\theta_M^N = 1$ iff $N \geq M$. The cached variables are now

$$\begin{aligned} u_j^{(k)} := & \frac{1}{8} \sum_{l=1}^{N-1-k} (1 - \delta_{k,l}) (F_{j+k}^{(l)} - F_j^{(l)}) (F_{j+k}^{(k+l)} - F_j^{(k+l)}) \\ & + \frac{1}{8} \sum_{l=k+1}^{N-1} (1 - \delta_{2k,l}) (F_{j+k}^{(l)} - F_j^{(l)}) (F_{j+k}^{(l-k)} - F_j^{(l-k)}) - \frac{1}{8} \sum_{l=1}^{k-1} (F_{j+k}^{(l)} - F_j^{(l)}) (F_{j+k}^{(k-l)} - F_j^{(k-l)}). \end{aligned} \quad (14b)$$

Again, these cached variables only involve values of $F_i^{(l)}$ that are not currently being optimized over and are thus static from the perspective of the k th-column optimization.

In this second-order approach, the objective function (14a) costs only $\mathcal{O}(N)$ to evaluate, while updating the cached variables costs $\mathcal{O}(N^2)$. Thus, the second-order approach is more efficient than the full-order approach (13) by $\mathcal{O}(N)$, at the cost of having some residual error (which itself is already better than the residual error in the first-order analytic solution).

If the optimization is performed using a gradient-based algorithm, the second-order approach also has the advantage of having a simple analytic gradient. The gradient of (14) in the direction of $F_i^{(k)}$ is given by

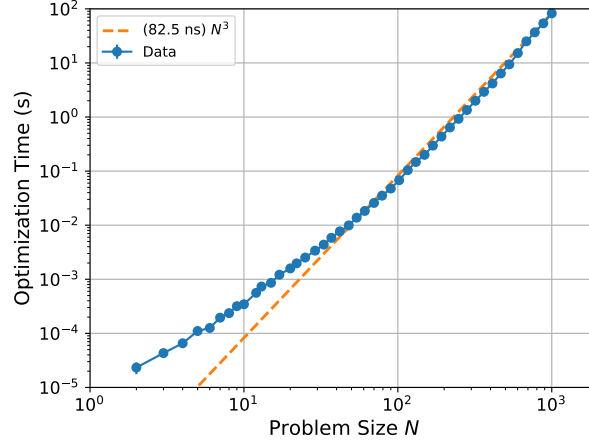
$$\sum_{j=1}^{N-k} v_j^{(k)} \left(\frac{\partial F_{j+k}^{(k)}}{\partial F_i^{(k)}} - \frac{\partial F_j^{(k)}}{\partial F_i^{(k)}} \right), \text{ where} \quad (15a)$$

$$v_j^{(k)} := \frac{1}{2} \left(\frac{\lambda_C C_j^{(k)}}{J_j^{(k)}} + \frac{1}{4} (F_{j+k}^{(k)} - F_j^{(k)}) \left[2 + \theta_{2k}^{N-1} (F_{j+k}^{(2k)} - F_j^{(2k)}) \right] + u_j^{(k)} \right) \left[2 + \theta_{2k}^{N-1} (F_{j+k}^{(2k)} - F_j^{(2k)}) \right]. \quad (15b)$$

The partial derivatives in (15a) are not explicitly evaluated because their values depend on the choice of symmetries in $F_j^{(k)}$; in an upper-triangular form, for example, the first term would be zero and the second a Kronecker delta; for a symmetry choice like in (10), the matrix of partial derivatives would be a sparse matrix. Thus, despite the sum in (15a), the evaluation of a single element of the gradient has $\mathcal{O}(1)$ cost, as long as the matrix of partial derivatives has $\mathcal{O}(1)$ nonzero entries per column (i.e., per fixed k).

In summary, provided that we choose a parametrization resulting in a sparse matrix of partial derivatives in $F_i^{(l)}$, the cost of optimizing the Floquet design parameters using a second-order approach is $\mathcal{O}(N \cdot (N^2 + MN))$, where M is the number of gradient and objective function evaluations needed to converge at each fixed k . The outside factor of N is needed because we visit each value of k (i.e., column) a constant number of times, while the factor of N^2 is the cost of the caching. If $M = \mathcal{O}(N)$, then the total cost is $\mathcal{O}(N^3)$.

To validate this analysis, we implemented and ran a program using the above optimization method for solving the second-order Floquet design problem. To test the scaling of the optimization procedure, we run it on random instances of SK7 across different problem sizes. The results of this numerical experiment are summarized in Sup. Figure 1, from which it is clear that 1) the design problem can be solved with only moderate processing time, even for a problem size of $N = 10^3$; and 2) the scaling at large N goes as $\mathcal{O}(N^3)$, as was argued in the discussion above.



Supplementary Figure 1. Time taken to solve Floquet design problem. Time taken to solve the Floquet design problem using a second-order, “column-by-column” approach as described in Sup. Note 1 C and implemented in Julia. The optimization time is the time required to compute the $F_j^{(k)}$ Floquet design parameters and is shown, for each problem size N , as an average over ten SK7 problem instances. For the procedure used to select the structure of the native couplings J_{ij} , see Sup. Note 5 A, and for details on the optimization routine, see Sup. Note 1 C 1. This calculation was performed with a single Julia process (i.e., using a single core) running on an Intel i7-5600U CPU at a nominal clock speed of 2.60 GHz. Note that error bars (standard error on the mean) are shown but are not visible on all points.

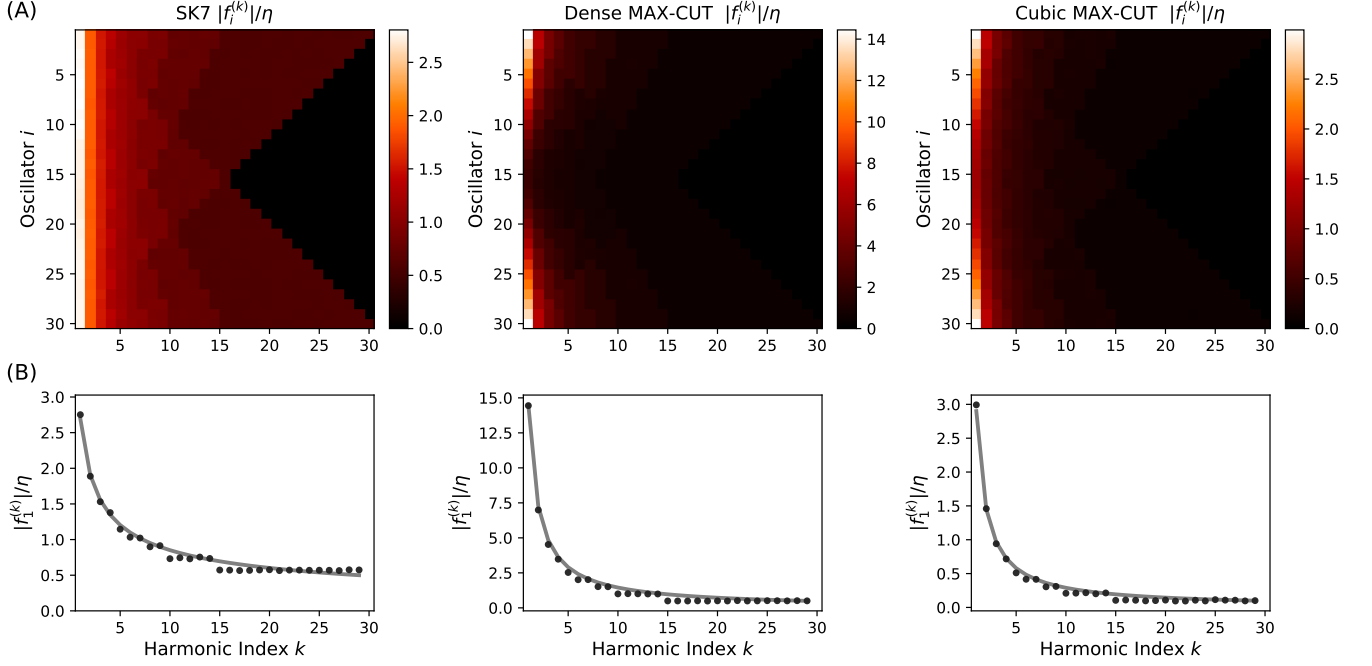
1. Additional details on numerical optimization routine

For the purposes of performing simulations and other numerical studies in this work, we implemented a straightforward version of the optimization routines described above. Our program is written in the Julia language (version 0.6.4), and the numerical optimization is performed by the Julia package Optim.jl (version 0.14.1) [3]. We mention here some additional details particular to our methodology, which may be relevant to the interpretation of our results.

We utilize the column-by-column approach as described in (13) or (14) (for a second-order Taylor approximation to the former), since it is more scalable than the joint approach described by (12). In this approach, there are $N - 1$ objective functions, one for each value of k in (13) and (14). In each of these objective functions, the optimization variables are $(F_1^{(k)}, \dots, F_{N-k}^{(k)})$; the other values of $F_j^{(k)}$ are determined by symmetry, using (10). Thus, if the Floquet design parameters $F_j^{(k)}$ were collected as a matrix, then the k th objective function involves optimizing over the k th “column” of that matrix. The goal is to iteratively optimize each of these columns, using the first-order analytic solution (10) as the starting point.

For each column being optimized, we make a call to the optimization routine provided by Optim.jl, using the conjugate gradient method with at most 250 iterations (i.e., steps) [3]. The value of the attained objective function minimum is then analyzed. If this minimum is below a certain tolerance level, then we deem that column to be “converged” and remove it from the set of columns that require optimizing further. This tolerance level is chosen for this work to be $(10^{-3}\eta)^2(N - k)$, where η is the dynamical-coupling parameter (see Sup. Note 1 D, 1 E, and 5 A). Interpretationally, this choice of the tolerance implies that the average element-wise residual error in solving the Floquet design equation (6) is approximately 10^{-3} , at least for the case of approximately uniform native couplings $J_{ij} \approx \lambda_J$. (However, we note that for the second-order optimizer, there is also a residual error due to the Taylor approximation as well, which increases with $\lambda_C/\lambda_J = \eta$ and can in fact dominate the residual error.)

Once every column to be optimized has been visited, the program then recomputes the objective function values for each column that has converged. Since the various columns are coupled, it is possible that the objective function value for, say, $k = 1$, previously below the tolerance, now exceeds the tolerance due to the optimization over, say, $k = 5$. Thus, if any column that has converged now exceeds the tolerance, it is added back into the set of columns requiring additional optimization. This new set of columns to optimize then becomes the starting point for another “pass”, and this whole procedure repeats until either 1) the number of such passes exceeds some maximum (set to 25 in this work), or if there are no longer any columns left to optimize. Furthermore, a column which fails to improve its objective function value by more than a fixed amount (set to 1 % in this work) on two consecutive passes is also removed (permanently) from the set of columns to optimize. Finally, at the end of the routine, if any of the objective function values are worse than those of the first-order analytic solution (10) (specifically, exceeding the latter by more than 1 % of the tolerance level), then the optimization results are rejected and the latter is returned.



Supplementary Figure 2. Structure of the Floquet design parameters. Magnitude of the Floquet design parameters $f_j^{(k)}$ in the first-order small-modulation approximation (see Sup. Note 1 B) and assuming a constant, uniform native couplings $J_{ij} = \lambda_J$, for a problem size of $N = 30$. Each point is computed as a mean over 5000 problem instances. (A): Heat map showing the mean magnitude as a function of oscillator index i and the harmonic index k . (B): The mean magnitude for the first oscillator, which is equivalent to the first row of each image in (A). The black dots are values produced from numerically averaging over many samples of problem instances while the gray solid line represents the analytic expressions given by (18).

D. Structure of the Floquet design parameters

Since the Floquet design parameters dictate the physical behavior of the oscillators undergoing phase modulation, it is useful to understand the general structure of $F_j^{(k)}$. Here, we provide some qualitative analysis which yields relationships that will be useful later on in the analysis of errors and scaling in hardware-specific contexts. More specifically, we analyze the small-modulation, first-order solution (10): in most of the situations we are concerned with, the numerical solutions do not drastically change the qualitative structure at first order.

The small-modulation solution (10) depends on both the target couplings C_{ij} as well as the native couplings J_{ij} . Let us denote by $f_j^{(k)}$ the first-order small-modulation solution to the design problem with J_{ij} taken to be a uniform and time-independent (i.e., $J_{ij}(t) = \lambda_J$ for $i \neq j$ and $\lambda_J \in \mathbb{R}^+$). This case anticipates approximations we later apply in which the hardware-provided native couplings are, in fact, quite close to this form (see Sup. Note 4).

We see that the dynamical-coupling ratio

$$\eta := \frac{\lambda_C}{\lambda_J} \quad (16)$$

naturally appears in the expression for (10), and it fully captures the dependence of $f_j^{(k)}$ on λ_C and λ_J . However, the utility of the ratio η extends beyond just analyzing $f_j^{(k)}$; as we will see, η plays a critical role in determining both the performance (see Sup. Note 1 E) as well as the scaling (see Sup. Note 5) of the dynamical-coupling scheme. Intuitively, it captures how much of the native coupling strength (represented by λ_J) is utilized to generate the effective target coupling strength (represented by λ_C).

We present in Sup. Figure 2 the average magnitude of $f_j^{(k)}$ for each of the representative problem classes. In Sup. Figure 2 A, we see that, quite generally, the oscillators at the two extremes of large and small detunings (i.e., nearer to the first and last rows) experience greater overall phase modulation than the oscillators closer to the middle. We also see that the magnitude of the phase modulation generally decreases as the harmonic of the modulation (i.e., the row index) increases. In addition, Sup. Figure 2 B, highlights the fact that the modulation strength for SK7 problems decays more slowly compared to the MAX-CUT problems.

The latter behavior can be understood by observing that (10) in this case simplifies to

$$f_1^{(k)} = \eta \sum_{m=1}^{\lfloor (N-1)/k \rfloor} C_{1+(m-1)k}^{(k)}. \quad (17)$$

Assuming that $\lfloor (N-1)/k \rfloor \approx (N-1)/k$ and making use of the statistical properties of each problem class, we can derive that

$$\text{mean } |f_1^{(k)}| \approx \begin{cases} \eta/2 \cdot (N-1)^{1/2} \cdot k^{-1/2} & \text{for SK7} \\ \eta/2 \cdot (N-1) \cdot k^{-1} & \text{for Dense MAX-CUT} \\ 3\eta \cdot (N-1)/N \cdot k^{-1} & \text{for Cubic MAX-CUT} \end{cases}. \quad (18)$$

We see that different problem classes have different scaling with respect to N and k . In the case of SK7, the target couplings C_{ij} have random alternations in sign, which causes cancellations in the sum of (17), so that the mean magnitude decays more slowly with modulation frequency. On the other hand, for sparse problem classes, there are asymptotically fewer terms in (17), leading to a different scaling. This functional form is shown in Sup. Figure 2B, which indicates good agreement with empirical results, on average.

E. Effective coupling error and the dynamical-coupling parameter

For the dynamical-coupling scheme to accurately approximate the target Hamiltonian, it is important for the effective couplings under time-averaging to approximate the target couplings accurately. Given a set of desired target couplings $\lambda_C C_{ij}$ and a (time-independent) set of native couplings J_{ij} , suppose we have solved the Floquet design equation (6) for the design parameters $F_j^{(k)}$ (using, for example, the methods of Sup. Note 1 C). We then define

$$C_{\text{eff}ij} := \int_0^{2\pi} \frac{J_{ij}}{\lambda_C} \cos \left[(i-j)t + \sum_{k=1}^{N-1} (F_i^{(k)} - F_j^{(k)}) \sin kt \right] \frac{dt}{2\pi} \quad (19)$$

to be the effective couplings produced by that set of $F_j^{(k)}$. Of course, if the Floquet design equation is solved well, C_{eff} should be close to C . Nevertheless, it is clear that C_{eff} can be different from C , especially if, for example, we use a ‘‘column-by-column’’ optimization approach or derive our objective functions from a small-modulation approximation.

For this work, we define the error of the effective couplings to be

$$\mathcal{E}_C := \max_{i \neq j} |C_{\text{eff}ij} - C_{ij}|. \quad (20)$$

This error metric of taking the maximum over all entries is conservative (one could choose to compute an element-wise average error instead, for example) and characterizes the largest element-wise deviation of the effective couplings from the target ones. We aim here to characterize the error $\mathcal{E}_C$ and to connect the scaling of the error with the dynamical-coupling parameter η .

In Sup. Note 1 B, we showed how the Floquet design equation can be simplified by an expansion in the modulation depth ζ . Intuitively, we expect the numerical algorithms presented in Sup. Note 1 C to perform better and produce lower error when the expansion parameter ζ is small. By examining (7), ζ is limited by the largest φ_{ij} . Furthermore, we have observed in Sup. Note 1 D that oscillators with the smallest and largest detunings ($i = 1, N$) are driven most strongly (so that φ_{1N} is largest), so we can reason that

$$\zeta \lesssim \max_t \varphi_{1N}(t) = \max_t \sum_{k=1}^{N-1} 2F_1^{(k)} \sin(k\Lambda t) \leq 2 \sum_{k=1}^{N-1} |F_1^{(k)}|, \quad (21)$$

where we have used $F_N^{(k)} = -F_1^{(k)}$ on the equality. We now further assume, as argued in Sup. Note 1 D, that the structure of $F_j^{(k)}$ is similar to the structure of $f_j^{(k)}$, derived in the first-order, constant- J_{ij} approximation. Then using (18), we arrive at the conclusion that ζ is, to a good approximation, *linearly proportional* to the dynamical-coupling parameter η , at least on average over problem instances. This in turn suggests that the error of the effective couplings reduces as η is made smaller. This is physically intuitive, as a small imposed ratio between the target coupling strength and native coupling strength implies more flexibility in engineering the system to approximate the desired target couplings.

Taken together, this discussion motivates the assertion that η is, in addition to Λ , one of the key dynamical-coupling parameters that determines the performance of the scheme. At the same time, given a fixed desired λ_C/χ in the target Hamiltonian, a small value of η implies a large value of $\lambda_J/\chi = (\lambda_C/\chi)/\eta$, which in turn implies a small value of the nonlinear rate χ ; this latter condition imposes a more stringent requirement on the coherence time of the system (see Sup. Note 5 C 3 for more details). Thus, a natural tradeoff is to choose the largest η that achieves an acceptable error.

It is reasonable to expect that, for any given problem instance, there is a precise value of η such that the Floquet design equation (6) fails to have a solution. While this “threshold” η varies across problem instances, we have found empirically (e.g., through numerical optimization) that, even on ensemble average over a problem class, there tends to be a sharp boundary between those values of η with acceptable errors after optimization and those for which the optimizer cannot improve the error relative to the first-order analytic solution (i.e., there is a dramatic change in the error once η exceeds a certain threshold value). This boundary value of η is a function of N , and we can understand its functional form by the following argument. We first substitute the analytic formulas for the mean magnitude of the Floquet design parameters $f_j^{(k)}$ from (18) into $F_j^{(k)}$ in (21). After some simplifying approximations (such as $N - 1 \approx N$), this results in

$$\zeta \lesssim \begin{cases} 2\eta \cdot N & \text{for SK7} \\ \eta \cdot N(1 + \log N) & \text{for Dense MAX-CUT} \\ 6\eta \cdot (1 + \log N) & \text{for Cubic MAX-CUT} \end{cases} \quad (22)$$

Heuristically, when ζ exceeds a certain modulation depth, the Floquet design equation tends to have no solution. If we denote this nominal value by $\zeta_{\max}$, then we can solve for η in the above to conclude that if we want a modulation depth smaller than $\zeta_{\max}$, we should take $\eta \leq \eta_{\text{bound}}$, given by

$$\eta_{\text{bound}} = \begin{cases} \zeta_{\max}/2 \cdot N^{-1} & \text{for SK7} \\ \zeta_{\max} \cdot (N(1 + \log N))^{-1} & \text{for Dense MAX-CUT} \\ \zeta_{\max}/6 \cdot (1 + \log N)^{-1} & \text{for Cubic MAX-CUT} \end{cases} \quad (23)$$

Sup. Figure 3 shows (in green) the range of values of η for which our numerical optimizer finds acceptable errors. Alongside this data (in pink) are the functional forms for η_{bound} derived above. We see that the functional forms are reproduced well across the different problem classes, while the offset can also be made to agree by setting $\zeta_{\max}$ to reasonable values (of order unity).

The upshot of this discussion is that the highest possible value of η as a function of N (and hence the best-case scaling of the dynamical-coupling scheme as a whole) is constrained by η_{bound} . In practice, however, finding the true solution of the Floquet design equation—or even being able to use the full-order numerical solver—is only possible for small problem sizes. As mentioned in Sup. Note 1 C, when the problem size is large, we generally turn to a second-order Taylor approximation approach for the optimizer. In this case, the error of the resulting solution is no longer sharp, as the quality of the second-order approximation (11) (which is used to construct the objective function) also depends on η (through ζ). In this case, we need to qualify our discussion by taking into account the choice of acceptable error. For simplicity, we consider $\mathcal{E}_C \leq 3\%$ as an acceptable error in this work.

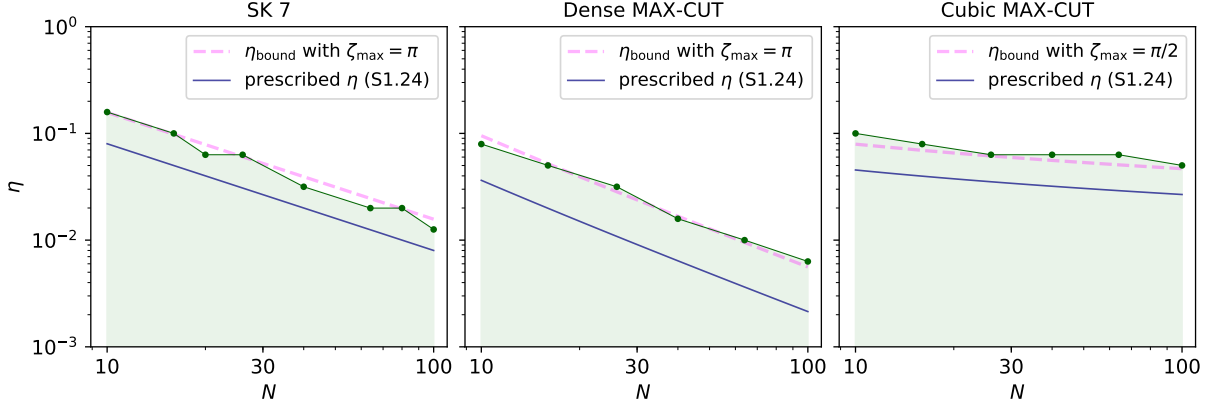
In Sup. Figure 4, we show the errors which are empirically attained by the second-order numerical optimizer. The green curve shows the values of η that achieve the acceptable error of 3%, below which the error is generally smaller. As expected, the 3% error boundary is qualitatively similar in shape to η_{bound} (cf., Sup. Figure 3), except for an overall offset due to the Taylor expansion. Thus, even though the second-order approach produces larger error overall, the acceptable-error cutoff for η scales in much the same way as the fundamental upper-bound η_{bound} discussed above.

Motivated by these observations, we choose η in this work by taking our theoretically derived fundamental bounds of η (i.e., η_{bound}) and multiplying them by an empirically determined multiplicative factor:

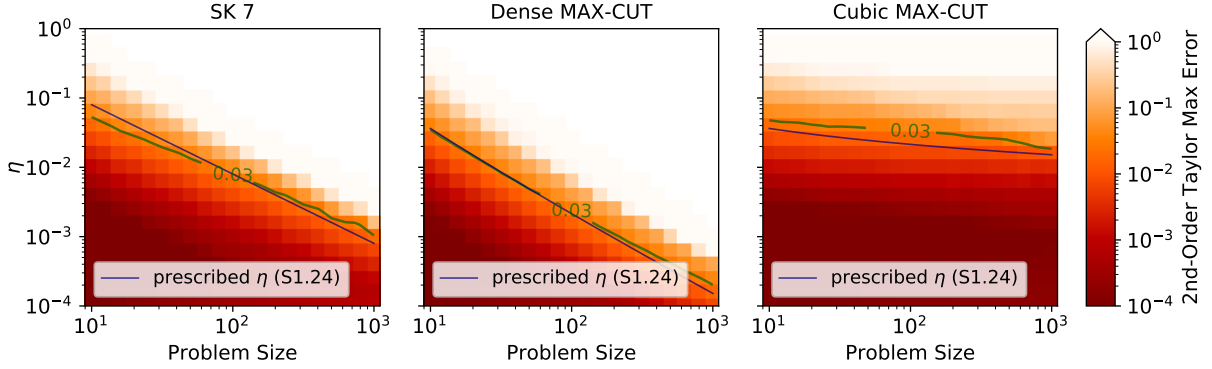
$$\eta = \begin{cases} 0.8 \cdot N^{-1} & \text{for SK7} \\ 1.2 \cdot N^{-1}(1 + \log N)^{-1} & \text{for Dense MAX-CUT} \\ 0.12 \cdot (1 + \log N)^{-1} & \text{for Cubic MAX-CUT} \end{cases} \quad (24)$$

This choice of η is plotted in blue in both Sup. Figures 3 and 4. Because of the scaling depicted in Sup. Figure 4, we have reasonable confidence that even if a second-order approach is used for solving the Floquet design problem, we are within a constant factor of the optimal scaling possible, as implied by Sup. Figure 3. As such, we use this prescription for η in all our simulations in this work as well as for the scaling analysis performed in Sup. Note 5 C 3.

As a technical point, we note that in Sup. Figure 4, there is a crossover between the blue and green lines, indicating that at smaller N , (24) is not quite enough to ensure $\mathcal{E}_C \leq 3\%$. However, we can circumvent this by simply using the full-order solver for $N \leq 100$ (which is efficient at small N), and only switch to the second-order approach for $N > 100$. Of course, this technicality does not affect any of our scaling arguments.



Supplementary Figure 3. Upper-boundary values of dynamical-coupling parameter. The region depicted in green indicates the range of the dynamical-coupling parameter η in which the numerical solution to the Floquet design equation—based on a full-order, “column-by-column” approach as described in Sup. Note 1 C—achieves low error $\mathcal{E}_C$, as a function of problem size N . Specifically, we choose $\mathcal{E}_C \leq 3\%$, although the qualitative features are insensitive to this choice due to the sharpness of the error transition. The error is computed as a mean over ten problem instances, and a mesh grid of 8 points (in N) and 31 points (in η) is used to generate the boundary curve. The pink lines show the heuristically derived functional forms for η_{bound} from (23). The blue lines indicate the choice of η as a function of N used in this work, as given by (24). For the choice of the native couplings J_{ij} , refer to Sup. Note 5 A; for details on the numerical optimization routine, see Sup. Note 1 C.



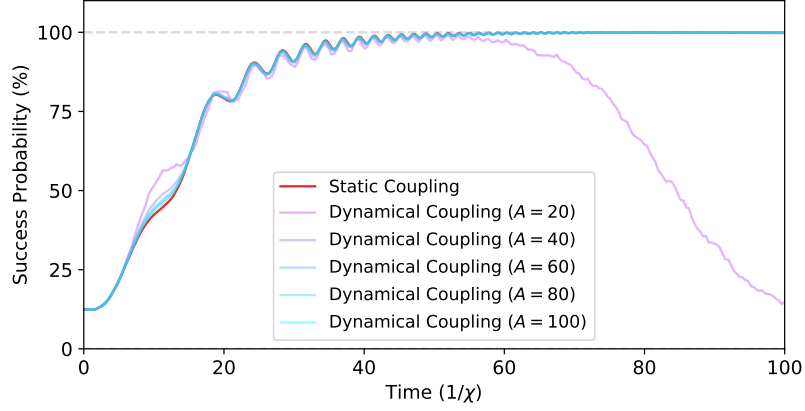
Supplementary Figure 4. Error scaling of the second-order numerical optimizer. Heat map of the error $\mathcal{E}_C$ resulting from the numerical solution to the Floquet design equation using a second-order, “column-by-column” approach as described in Sup. Note 1 C, plotted as a function of the dynamical-coupling parameter η and problem size N . The level curve for an acceptable error of $\mathcal{E}_C = 3\%$ is shown in green. The error is computed as a mean over 100 problem instances for each point. The blue lines indicate the choice of η as a function of N used in this work, as given by (24). For the choice of the native couplings J_{ij} , refer to Sup. Note 5 A; for details on the numerical optimization routine, see Sup. Note 1 C. Note that to save on computation, we omit the region where η has safely exceeded η_{bound} (namely, the upper-right triangular region in the left and center plots, around where the colormap saturates); by extrapolation, we take the error in this region to exceed 1.

F. Requirements on Floquet frequency

Finally, we discuss the Floquet frequency Λ , which sets the periodicity of used in the Floquet approximation to be $2\pi/\Lambda$. As mentioned above, the natively-coupling Hamiltonian (1) is approximated by the time-average effective Hamiltonian (3) only if Λ is much greater than all other system rates. If this assumption is not met, then it is possible that the physics of the dynamically-coupled system fails to approximate that of the target statically-coupled system, despite the Floquet equation (6) being met (i.e., despite $C_{\text{eff}ij} \approx C_{ij}$). More precisely, we require

$$\Lambda = A \times \max(\text{all system rates}), \quad (25)$$

where $A \gg 1$ is a large multiplicative factor. As will be discussed in later sections—see especially Sup. Note 2, 4, and 5 C 3 (and as presented in the main manuscript), the most relevant rates for our system are χ , δ_i , r_{max} , and λ_J .



Supplementary Figure 5. Timescale requirements for Floquet approximation. Comparison between the behavior of the statically-coupled target system (red, in back), vs. that of the corresponding dynamically-coupled system using a range value for the Floquet frequency Λ (parameterized by the factor A , as defined in (25)). The metric plotted here is the success probability (see Sup. Note 6 A) on a particular problem instance (specifically, the instance used for Sup. Figure 2 B in the main manuscript). Here, the physical model for the oscillators is a bus-coupled system of JPOs (see Sup. Note 3), under appropriate approximations (see Sup. Note 4) and with parameters chosen according to Sup. Note 5 A. For quantum simulation methods, see Sup. Note 6 A.

In order to determine the value of A that suffices, we simulate both the target statically-coupled system and the corresponding dynamically-coupled system using different values of A and examine the respective evolution of a key performance metric, the success probability (see Sup. Note 6 A). In Sup. Figure 5, we show one example of how the dynamically-coupled system progressively matches the evolution of the target statically-coupled system (at least, on timescales greater than $2\pi/\Lambda$) as A is increased. For this particular problem instance and system parameters, we see that the value of A should be larger than 40 for good correspondence.

From such empirical observations, we conservatively use in this work a value of $A = 100$ for quantum simulations. For simulations of the classical EOMs, similar empirical investigations show that a larger value of $A = 200$ is needed for comparable convergence.

Supplementary Note 2. Quantum annealing with Kerr parametric oscillators

In this section, we review some operational principles for solving Ising problems with networks of Kerr parametric oscillators (KPOs) via quantum annealing. A number of approaches towards this end have been proposed [1, 4, 5], but for simplicity, we focus on the directly- and statically-coupled model used in Ref. [1], as this is the model that our dynamical-coupling scheme aims to approximate.

A generic KPO network is represented by a Hamiltonian of the general form

$$-\sum_i \delta_i \hat{a}_i^\dagger \hat{a}_i - \sum_i \frac{\chi}{2} \hat{a}_i^{\dagger 2} \hat{a}_i^2 + \sum_i \left(\frac{r}{2} \hat{a}_i^2 + \text{H.c.} \right) - \lambda_C \sum_{i \neq j} C_{ij} \hat{a}_i^\dagger \hat{a}_j, \quad (26)$$

where $\chi > 0$ is the Kerr nonlinear rate, r is the amplitude of the parametric pump drive, δ_i is the detuning of the i th oscillator from the half-harmonic of the pump, $\lambda_C C_{ij}$ is the strength of the coupling between oscillator i and j . Generally, the symmetric coupling matrix C_{ij} is chosen to be that of a (bias-free) Ising problem whose ground state is desired. Finally, although implicit in (26), all these parameters can potentially be time-dependent, varying as needed by the annealing procedure.

In the absence of coupling and detuning, the eigenstates of each KPO are the 0-phase and π -phase coherent states $|+\sqrt{r/\chi}\rangle$ and $|-\sqrt{r/\chi}\rangle$, which can be utilized in a quantum annealer to encode, respectively, the spin-up and spin-down configurations of an Ising spin system. In a quantum annealer based on this encoding, the state of the KPO network after the annealing procedure is ideally the multimode Schrödinger cat state

$$\bigotimes_{i=1}^N |\sigma_i \sqrt{r/\chi}\rangle + \bigotimes_{i=1}^N |-\sigma_i \sqrt{r/\chi}\rangle, \quad (27)$$

where σ is the ground-state configuration of the Ising problem of interest; and it comprises $\sigma = (\sigma_1, \dots, \sigma_N)$ with $\sigma_i \in \{-1, 1\}$. Given this state, measurement of the sign of the in-phase field amplitudes of the KPOs produces solutions to the Ising problem.

We can consider two approaches in which such a KPO network can be used for performing quantum annealing and obtaining the final state (27). One way is to take a purely qubit-based approach [6], as the KPO has been shown to exhibit a qubit subspace (spanned by $|\pm\sqrt{r/\chi}\rangle$) in which (26) can be used to realize universal control [7, 8]. In this approach, the individual KPOs are each first prepared in the Pauli- X +1-eigenstate, which corresponds to the cat state $|\sqrt{r/\chi}\rangle + |-\sqrt{r/\chi}\rangle$; as described in [8], this can be done by engineering the evolution of the pump parameter r appropriately. After the preparation step, the annealing procedure then requires realizing and controlling two Hamiltonian terms in the qubit basis. The first is the so-called driver Hamiltonian $\sum_i \hat{X}_i$, which can be realized in the qubit subspace by proper choice of the detunings δ_i [7, 8]. The second is the problem Hamiltonian $\sum_{i \neq j} C_{ij} \hat{Z}_i \hat{Z}_j$, which is realized in the qubit subspace by the coupling term in (26). The key point here is that the state (27)—when interpreted in the qubit subspace—is the ground state of this latter Hamiltonian, so that the end result is the state (27), provided the assumptions of the quantum adiabatic theorem are met (i.e., in the context of standard adiabatic quantum computation). In this approach, the pump is held fixed to stabilize the two coherent states $|\pm\sqrt{r/\chi}\rangle$ and ensure that the system remains in the qubit subspace, while the coupling term is gradually turned on by adiabatically time-varying λ_C .

For our work, however, we focus on a more recent continuous-variable approach, as exemplified by Ref. [1] but also utilized in Refs. [4, 5]. This approach has the advantage of possessing a straightforward classical limit, allowing us to numerically study scaling to larger problem sizes. In contrast to the purely qubit-based approach, the initial state of the quantum annealer in this continuous-variable approach is the vacuum state. Then, the quantum annealing procedure is realized by time-varying the parametric pump drive according to some function $r(t)$, where $r(0) = 0$. A useful choice for $r(t)$ takes the form of a linear ramp with slope $r_{\max}/T_{\text{ramp}}$ clamped at $r_{\max}$:

$$r(t) = r_{\max} \min(t/T_{\text{ramp}}, 1). \quad (28)$$

The detunings δ_i are static but independently tuned for each oscillator such that the highest-eigenvalue eigenstate of (26) at $r = 0$ is the vacuum, and one satisfactory choice is [1]

$$\delta_i = \lambda_C \sum_j |C_{ij}|. \quad (29)$$

This choice ensures that the initial vacuum state of the oscillator network is adiabatically connected to the multimode cat state (27) at the end of the computation, thus implementing quantum annealing [1]. In this approach, the coupling

term $\lambda_C \sum_{i,j} C_{ij} \hat{a}_i^\dagger \hat{a}_j$ is also held static throughout the procedure, with some suitable choice of the problem-dependent strength parameter λ_C ; we discuss this tuning parameter further in Sup. Note 2 A below.

Of course, these networks of KPOs are not only described by the Hamiltonian (26), as they are also coupled to the environment. Following the approach in [4, 5], we model this dissipation due to the environment with Lindblad operators [9]. In particular, we treat the photon loss for the i th oscillator with the Lindblad operator

$$\hat{L}_i = \sqrt{\kappa} \hat{a}_i \quad (30)$$

where κ is the cavity-photon decay rate associated with this dissipation process.

A. Parameter considerations for annealing

The performance of a continuous-variable quantum annealer can be quite sensitive to the choice of parameters and annealing schedules used in the scheme. Focusing on the continuous-variable annealing scheme with a clamped-linear pump schedule as in Ref. [1], we discuss in this subsection the prescriptions we follow for selecting the ramp time T_{ramp} , the maximum pump rate r_{max} , and the problem-strength parameter λ_C . These prescriptions are then directly mapped onto the corresponding parameters in our dynamical-coupling scheme.

Among the three parameters, λ_C is the most subtle, as its optimal choice depends on the problem size N as well as the problem structure. To first order, Ref. [1] shows that the KPO system is an effective quantum annealer under the condition that

$$\left| -\sum_i \frac{\chi}{2} \hat{a}_i^{\dagger 2} \hat{a}_i^2 + \sum_i \left(\frac{r(t)}{2} \hat{a}_i^2 + \text{H.c.} \right) \right| \gg \left| \sum_i \delta_i \hat{a}_i^\dagger \hat{a}_i + \lambda_C \sum_{i \neq j} C_{ij} \hat{a}_i^\dagger \hat{a}_j \right|. \quad (31)$$

Thus, in order for the annealer to be effective, we should scale the value of λ_C to cancel out the scaling in the sum over C_{ij} , to ensure the right-hand side remains bounded as required.

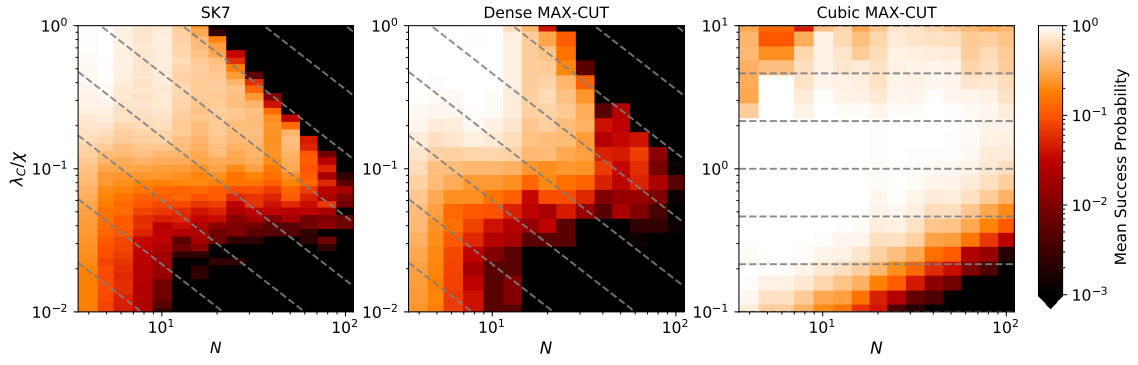
We present in Sup. Figure 6 the success probability, computed via simulations of the classical EOMs that are described in Sup. Note 6 B, of various problem classes as we vary λ_C and N . For simplicity the loss is set to zero $\kappa = 0$ for these simulations. The result shows that dense problem classes (SK7, Dense MAX-CUT) require $\lambda_C/\chi \propto 1/N$, while sparse problem classes (Cubic MAX-CUT) require a constant λ_C . This numerical investigation is consistent with (31) in that the optimal choice for the problem-strength parameter is $\lambda_C/\chi \sim r_{\text{max}} (\sum_{ij} |C_{ij}|)^{-1}$. Based on these results, unless stated otherwise, we recommend the following forms for λ_C :

$$\frac{\lambda_C}{\chi} = \begin{cases} 4/N & \text{for SK7} \\ 8/N & \text{for Dense MAX-CUT} \\ 2 & \text{for Cubic MAX-CUT} \end{cases} \quad (32)$$

In this work, we also analyze the performance of the statically-coupled system with dissipation to ensure that the dynamical-coupling scheme is (at least comparably) robust to effects of linear loss. The presence of dissipation affects the choice for the parameters T_{ramp} , r_{max} and λ_C that optimize success probability. To investigate these effects, we perform quantum simulations of the statically-coupled system and present them in Sup. Figure 7, which shows how the success probability of solving $N = 4$ SK7 problems varies for different settings of the parameters λ_C , r_{max} and T_{ramp} , across the three different loss rates $\kappa/\chi \in \{0, 2 \times 10^{-2}, 2 \times 10^{-1}\}$, which are presented along each row respectively.

First, we see that the optimal choice of T_{ramp} varies strongly with κ . In the absence of loss, success probability monotonically improves as T_{ramp} is increased, since the annealing procedure becomes more adiabatic. However, with finite loss, larger T_{ramp} causes a decrease in the success probability, as the number of photon emission events accumulate, causing decoherence. Based on the results of Sup. Figure 7, the ramp time should be set to $T_{\text{ramp}} = \min(2/\kappa, 100/\chi)$ to maximize success probability, corresponding to order unity number of photon emission events per oscillator in the duration of the computation [4] (and where $100/\chi$ gives acceptable adiabaticity in the lossless case).

Next, r_{max} determines the amplitude of the coherent states that constitute the multimode cat at the end of the computation. Thus, it should be sufficiently large to ensure that the two constituent coherent states are distinguishable (i.e., nearly orthogonal). On the other hand, when there is loss, r_{max} should also not be too large, as the number of photon emission events increases with the cat-state amplitude, reducing success probability. The numerical simulations suggest that $r_{\text{max}} = 5$ achieves reasonable success probabilities even in the presence of dissipation. Finally, we note that there is not very strong dependence of λ_C with dissipation, as we find that the value given by (32) results in reasonably good success probabilities even in the presence of photon loss, at least for $N = 4$.



Supplementary Figure 6. Dependence of KPO network annealer on problem-strength parameter. Success probability of the classical EOMs for statically-coupled KPO networks (105) as a function of the problem-strength parameter λ_C/χ and problem size N , for the three problem classes. The success probability is computed as a mean over at least ten problem instances for each class. The gray dashed lines show curves that are linearly proportional to the functional forms given in (32). For the approach to simulating classical EOMs, see Sup. Note 6 B.

Of course, in addition to these qualitative observations about the operating point of the statically-coupled system, one can perform additional fine-tuning. For the quantum simulations in this work, we have chosen specific values of the parameters using Sup. Figure 7 to maximize success probabilities; see Sup. Table 1.

Supplementary Note 3. Bus-mediated JPO networks

A. Josephson parametric oscillators

A Josephson parametric oscillator (JPO) is a Kerr resonator with an additional two-photon drive. JPOs can be implemented in different ways, such as with a SNAIL parametric amplifier [10] or by parametric modulation of a transmon oscillator's resonance frequency [4, 8]. For this work, we focus on the latter implementation, where the magnetic flux Φ (henceforth normalized to the flux quantum $\Phi_0 := h/2e$) applied to the DC SQUID is modulated instead of being held static. The system is described by the standard transmon Hamiltonian [11]

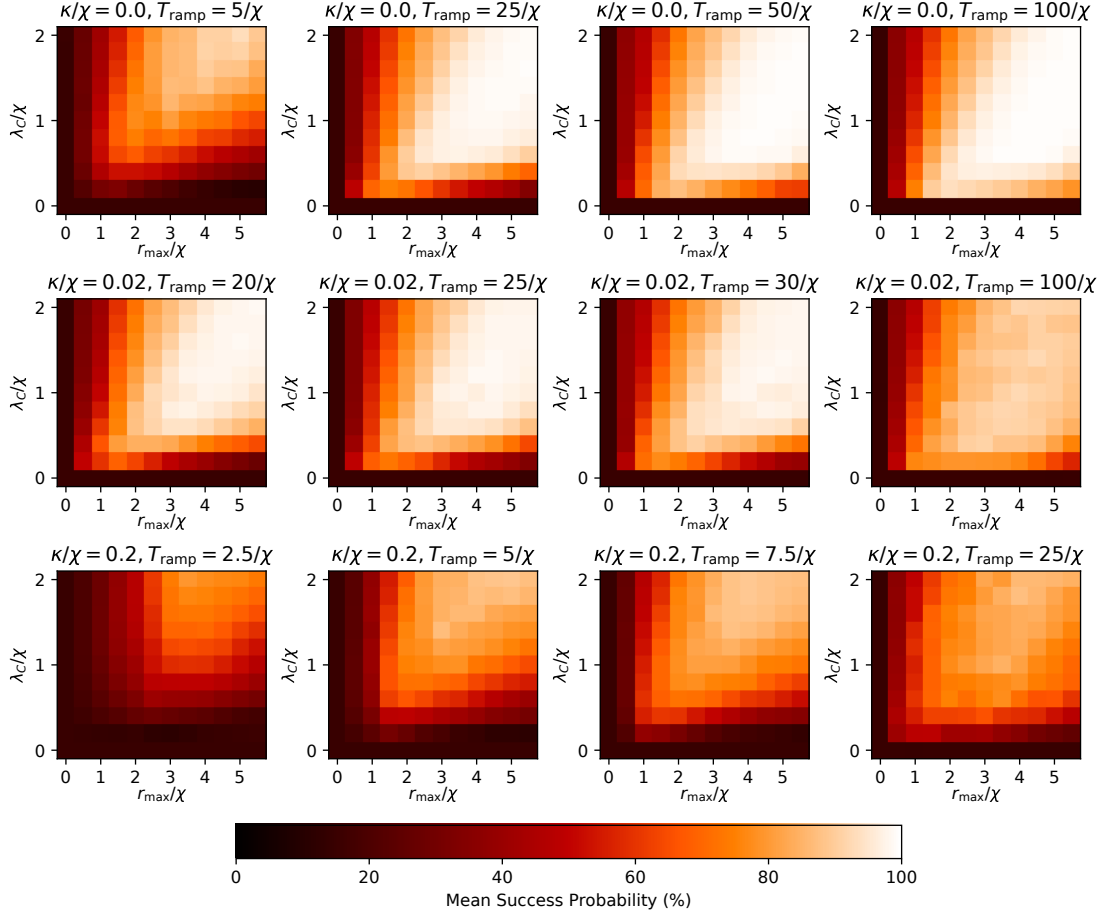
$$4E_C\hat{q}^2 - 2E_J \cos(\Phi/2) \cos \hat{\theta}, \quad (33)$$

where $\hat{q}, \hat{\theta}$ are canonically conjugate variables representing the charge on the capacitor and the phase difference of the two junctions, respectively, and E_C and E_J are the charging and Josephson energies, respectively.

We posit that the flux $\Phi(t)$ has a time-dependence which can be separated into DC and AC components, as

$$\Phi(t)/2 := \Phi_{\text{dc}} + \Phi_{\text{ac}}(t). \quad (34)$$

Operationally, Φ_{ac} is used for both parametric driving (at high frequencies) and for detuning modulation (at lower frequencies) to achieve dynamical coupling while Φ_{dc} is used to set the operating frequency of each JPO in the absence



Supplementary Figure 7. Dependence of KPO network annealer on pump rate and ramp time. Success probability of the statically-coupled KPO quantum annealer at solving SK7 problems with problem size $N = 4$, as functions of cavity-photon decay rate κ/χ (rows) and the linear ramp time $T_{\text{ramp}}\chi$ (columns), as well as the maximum pump rate r_{max}/χ and problem-strength parameter λ_c/χ . The success probability is computed as a mean over ten problem instances. For quantum simulation approach, see Sup. Note 6 A.

of the modulation. Assuming the modulation amplitude is small (i.e., $|\Phi_{\text{ac}}(t)| \ll \Phi_{\text{dc}}$), we can expand [4]

$$\begin{aligned}\cos(\Phi(t)/2) &= \cos[\Phi_{\text{dc}}] \cos[\Phi_{\text{ac}}(t)] - \sin[\Phi_{\text{dc}}] \sin[\Phi_{\text{ac}}(t)] \\ &\approx \cos[\Phi_{\text{dc}}] - \sin[\Phi_{\text{dc}}] \Phi_{\text{ac}}(t).\end{aligned}\quad (35)$$

In the transmon limit of $E_J \gg E_C$ [12], we can also expand (33) in $\hat{\theta}$. Dropping the zeroth-order (constant) term and inserting (34) after applying the approximation (35), we have, to fourth order in $\hat{\theta}$, the Hamiltonian

$$4E_C \hat{q}^2 + E_J \cos(\Phi_{\text{dc}}) \hat{\theta}^2 - E_J \sin(\Phi_{\text{dc}}) \Phi_{\text{ac}}(t) \hat{\theta}^2 - \frac{E_J}{12} \cos \Phi_{\text{dc}} \hat{\theta}^4. \quad (36)$$

We next introduce bosonic operators $\hat{a}$ and $\hat{a}^\dagger$ (satisfying $[\hat{a}, \hat{a}^\dagger] = 1$) via the canonical transformations [4]

$$\hat{\theta} = \left[\frac{E_C}{E_J \cos(\Phi_{\text{dc}})} \right]^{1/4} (\hat{a} + \hat{a}^\dagger), \quad \hat{q} = -i \left[\frac{E_J \cos(\Phi_{\text{dc}})}{16E_C} \right]^{1/4} (\hat{a} - \hat{a}^\dagger). \quad (37)$$

Inserting (37) into (36), we arrive at the Hamiltonian

$$4\sqrt{E_C E_J \cos(\Phi_{\text{dc}})} \hat{a}^\dagger \hat{a} - \frac{E_C}{12} (\hat{a} + \hat{a}^\dagger)^4 - \sqrt{\frac{E_C E_J}{\cos(\Phi_{\text{dc}})}} \sin(\Phi_{\text{dc}}) \Phi_{\text{ac}}(t) (\hat{a} + \hat{a}^\dagger)^2. \quad (38)$$

Finally, after normal ordering and applying the rotating wave approximation [13] to cut off frequencies much higher than twice the fundamental (ω_{dc} in (40)), we have the dynamically-controlled JPO Hamiltonian

$$\hat{H}_{\text{JPO}}(t) := \omega(t) \hat{a}^\dagger \hat{a} - \frac{\chi}{2} \hat{a}^{\dagger 2} \hat{a}^2 + \left(\frac{\omega_{\text{ac}}(t)}{2} \hat{a}^2 + \text{H.c.} \right) \quad (39)$$

where we have defined the nonlinear Kerr coefficient $\chi := E_C$ and the instantaneous frequency $\omega(t) := \omega_{\text{dc}} + \omega_{\text{ac}}(t)$, which has DC and AC components given by

$$\omega_{\text{dc}} := 4\sqrt{E_C E_J \cos(\Phi_{\text{dc}})} - E_C, \quad \omega_{\text{ac}}(t) := -2\sqrt{\frac{E_C E_J}{\cos(\Phi_{\text{dc}})}} \sin(\Phi_{\text{dc}}) \Phi_{\text{ac}}(t). \quad (40)$$

These relations indicate that, for appropriate choices of E_C and E_J , both the DC and AC components of the instantaneous frequency can be fully controlled in a JPO given appropriate control over Φ_{dc} and $\Phi_{\text{ac}}(t)$.

B. Bus-mediated native couplings

To couple together JPOs that are individually described by (39), we utilize an architecture involving a common bus resonator. We consider an excitation-exchange interaction between the JPOs and the bus, and the linear part of system Hamiltonian (i.e., omitting the individual JPO's parametric and Kerr terms) is [14]

$$\hat{H}_{\text{coupling}} := \sum_i \omega_i \hat{a}_i^\dagger \hat{a}_i + \omega_{\text{bus}} \hat{b}^\dagger \hat{b} + \sum_i g_i \left(\hat{b}^\dagger \hat{a}_i + \hat{b} \hat{a}_i^\dagger \right), \quad (41)$$

where ω_{bus} is the frequency of the bus resonator and g_i is the coupling rate of the i th JPO to the bus. As in (39), the instantaneous frequency ω_i is in general a time-dependent parameter.

We start by introducing the bus-oscillator detunings

$$\Delta_i := \omega_{\text{bus}} - \omega_i. \quad (42)$$

Next, we make the assumption that, for $\varepsilon \ll 1$

$$g_i/\Delta_i \sim \mathcal{O}(\varepsilon), \quad \text{and} \quad \chi_i/\Delta_i \sim \omega_{\text{ac}i}/\Delta_i \sim \mathcal{O}(\varepsilon^2), \quad (43)$$

for all i and at all times. Provided these conditions are met, we can perform a Schrieffer-Wolff expansion [12] in the rotating frame of the bus, to second order in ε . The result of that expansion produces the coupling Hamiltonian

$$\hat{H}_{\text{coupling}} \approx \sum_i \left(\omega_i - \frac{g_i^2}{\Delta_i} \right) \hat{a}_i^\dagger \hat{a}_i - \sum_{i \neq j} \frac{g_i g_j}{2} \left(\frac{1}{\Delta_i} + \frac{1}{\Delta_j} \right) \hat{a}_i^\dagger \hat{a}_j, \quad (44)$$

where we have dropped the bus mode operators $\hat{b}$ in (44) since the bus will no longer be populated in this limit. Note that due to the Schrieffer-Wolff expansion, this Hamiltonian is only approximately in lab frame, as there is an $\mathcal{O}(\varepsilon)$ ‘‘polariton angle’’ between the JPO and bus modes: the $\hat{a}_i$ operators used in (44) are hybridized with $\hat{b}$ to $\mathcal{O}(\varepsilon)$. We make the standard assumption in circuit QED that this has negligible effect [12].

C. JPO control problem

Taking the results of the two above subsections together, the (approximately) lab-frame Hamiltonian of a system of JPOs with bus-mediated couplings is

$$\sum_i \left(\omega_i(t) - \frac{g_i^2}{\Delta_i(t)} \right) \hat{a}_i^\dagger \hat{a}_i - \frac{\chi_i}{2} \hat{a}_i^{\dagger 2} \hat{a}_i^2 + \left(\frac{\omega_{ac i}(t)}{2} \hat{a}^2 + \text{H.c.} \right) - \sum_{j \neq i} \frac{g_i g_j}{2} \left(\frac{1}{\Delta_i(t)} + \frac{1}{\Delta_j(t)} \right) \hat{a}_i^\dagger \hat{a}_j. \quad (45)$$

Our goal is to choose the (time-dependent) parameters such that this Hamiltonian approximates the generic KPO quantum annealing Hamiltonian (26). We first go into a rotating frame in which the i th JPO oscillates at a constant frequency of $-\delta_i$. This produces the Hamiltonian

$$- \sum_i \left(\delta_i \hat{a}_i^\dagger \hat{a}_i + \frac{\chi_i}{2} \hat{a}_i^{\dagger 2} \hat{a}_i^2 \right) + \hat{H}_p + \hat{H}_c,$$

where the pump and coupling terms are, respectively,

$$\hat{H}_p := \sum_i \frac{\omega_{ac i}(t)}{2} \exp \left[-2i \int_0^t \left(\omega_i(t') - \frac{g_i^2}{\Delta_i(t')} + \delta_i \right) dt' \right] \hat{a}_i^2 + \text{H.c.} \quad (46a)$$

$$\hat{H}_c := - \sum_{i \neq j} \frac{g_i g_j}{2} \left(\frac{1}{\Delta_i(t)} + \frac{1}{\Delta_j(t)} \right) \exp \left[-i \int_0^t \left(\omega_j(t') - \frac{g_j^2}{\Delta_j(t')} + \delta_j - \omega_i(t') + \frac{g_i^2}{\Delta_i(t')} - \delta_i \right) dt' \right] \hat{a}_i^\dagger \hat{a}_j. \quad (46b)$$

Comparing this Hamiltonian with (26), we see that the design problem is to obtain

$$\hat{H}_p \approx \sum_i \frac{r(t)}{2} \hat{a}_i^2 + \text{H.c.} \quad (47a)$$

$$\hat{H}_c \approx -\lambda_C \sum_{i \neq j} C_{ij} \hat{a}_i^\dagger \hat{a}_j. \quad (47b)$$

We see that $\hat{H}_c$ is already written in the form of (1), so in principle, we can use our approach to dynamical coupling from Sup. Note 1 to parameterize $\omega_i(t)$ in terms of the Floquet design parameters $F_i^{(k)}$ and thus satisfy (47b) by solving the Floquet design problem for $F_j^{(k)}$.

Unfortunately, for the specific case of JPOs, there is an additional complication because that same control signal ω_i is already used for performing the parametric pumping of the system: this is encapsulated by (47a) and (46a), which directly depends on the AC part of $\omega_i(t)$. We emphasize that this additional complication is due to the hardware constraints of JPOs and is not inherent to the design of dynamical couplings. (For example, in optical implementations of KPOs, a separate pump field is responsible for producing the parametric drive, opening a new degree of freedom for solving the design problem.)

To help get around this JPO-specific complication, we assume that the AC part of the instantaneous frequency has two “bands” of frequency content:

$$\omega_{ac}(t) = \omega_{\text{slow}}(t) + \omega_{\text{fast}}(t), \quad (48)$$

where, roughly speaking, ω_{slow} consists of frequency content near DC (the “coupling band”), while ω_{fast} consists of frequency content near $2\omega_{\text{dc}}$ (the “pumping band”). We now proceed to solve the problem by assuming that the following two replacements can be made in (46), at least in the context of trying to achieve the approximate equalities (47):

$$\frac{1}{\Delta_i(t)} \mapsto \frac{1}{\omega_{\text{bus}} - (\omega_i^{\text{dc}} + \omega_i^{\text{slow}}(t))} \quad (49a)$$

$$\exp \left[i \int_0^t \omega_i(t') dt' \right] \mapsto \exp \left[i \int_0^t (\omega_i^{\text{dc}} + \omega_i^{\text{slow}}(t')) dt' \right]. \quad (49b)$$

This trick allows us to effectively decouple the pumping and coupling problems, at the cost of potential “crosstalk” between the two Hamiltonian terms once we insert our solution for $\omega_i(t)$.

We first use the approach of dynamical coupling to solve the design problem of (47b). To match the abstract form of (1), we define the native coupling matrix and phase functions

$$J_{ij}(t) := \frac{g_i g_j}{2} \left(\frac{1}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t)} + \frac{1}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t)} \right) \quad (50a)$$

$$\phi_i(t) := \int_0^t \left(\omega_{\text{dc}i} + \omega_{\text{slow}i}(t') - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t')} + \delta_i \right) dt'. \quad (50b)$$

These definitions identify for us, in the two-band approximation, the manifestation of the native-coupling matrix $J_{ij}(t)$ and oscillator phases $\phi_i(t)$ in the context of the bus-coupled JPO network. Thus, following (4), we want to parameterize $\omega_{\text{dc}i}$ and $\omega_{\text{slow}i}(t)$ (on the left) in terms of the Floquet design parameters (on the right) according to

$$\int_0^t \left(\omega_{\text{dc}i} + \omega_{\text{slow}i}(t') - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t')} + \delta_i \right) dt' = \omega_0 t - (i-1)\Lambda t - \sum_{k=1}^{N-1} F_i^{(k)} \sin(k\Lambda t), \quad (51)$$

where the scalars $F_i^{(k)}$ are the solution to the Floquet design problem (6). Provided that we know the values of $F_j^{(k)}$ required, (51) is a (hardware) design problem for the control variables $\omega_{\text{dc}i}$ and $\omega_{\text{slow}i}(t)$. Differentiating both sides of (51),

$$\omega_{\text{dc}i} + \omega_{\text{slow}i}(t) - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t)} + \delta_i = \omega_0 - (i-1)\Lambda - \sum_{k=1}^{N-1} F_i^{(k)} \cos(k\Lambda t) k\Lambda. \quad (52)$$

We then split this equation into time-independent (DC) and time-dependent (slow) parts:

$$\omega_{\text{dc}i} - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i}} + \delta_i = \omega_0 - (i-1)\Lambda \quad (53a)$$

$$\omega_{\text{slow}i}(t) + \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i}} - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t)} = - \sum_{k=1}^{N-1} F_i^{(k)} \cos(k\Lambda t) k\Lambda. \quad (53b)$$

The first equation can be converted into a quadratic equation for $\omega_{\text{dc}i}$ and solved, as can the second equation for $\omega_{\text{slow}i}(t)$ after given $\omega_{\text{dc}i}$. (See subsection below for more details.)

As an aside, if $\delta_i \ll \Lambda$ and $g_i^2/\Delta_i(t) \ll \Lambda$ (which is necessary to satisfy the Schrieffer-Wolff conditions anyway), these equations (53) can be approximated by

$$\omega_{\text{dc}i} \approx \omega_0 - (i-1)\Lambda \quad (54a)$$

$$\omega_{\text{slow}i}(t) \approx - \sum_{k=1}^{N-1} F_i^{(k)} \cos(k\Lambda t) k\Lambda, \quad (54b)$$

which shows that this hardware manifestation supports our claim in Sup. Note 1 that the mean oscillator frequencies are approximately evenly spaced. It also shows that the magnitudes of the slow component of the instantaneous frequencies directly correlate with the magnitude of $F_i^{(k)}$ and is intuitively given by the derivative of the instantaneous phase modulation $\dot{\phi}_i(t)$.

It is also worth noting that technically, the determination of the Floquet design parameters $F_i^{(k)}$ requires knowledge of the matrix $J_{ij}(t)$, and in this system, $J_{ij}(t)$ also depends on the control variables $\omega_i(t)$, as shown by (50a). Thus, the whole problem is, in principle, highly coupled, and we cannot assume we have the values of $F_j^{(k)}$ when trying to solve (53)! A full numerical technique could iteratively solve the Floquet design problem (6) by proposing values for $F_i^{(k)}$, solving (53) for those proposed values, inserting the resulting value for $J_{ij}(t)$ into (5b), and repeating until converged. Alternatively, it is possible to approximate J_{ij} using an approximate solution to (53) that is *independent* of $F_i^{(k)}$, and this latter approach will be discussed in Sup. Note 4.

Finally, turning to the pumping design problem in (47a), the two-band approach says that we want

$$(\omega_{\text{slow}i}(t) + \omega_{\text{fast}i}(t)) \exp \left[-2i \int_0^t \left(\omega_{\text{dc}i} + \omega_{\text{slow}i}(t') - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc}i} - \omega_{\text{slow}i}(t')} + \delta_i \right) dt' \right] \approx r(t). \quad (55)$$

We note that the first term of this equation (proportional to $\omega_{\text{slow}i}(t)$) produces fast oscillations in the pumping band, whereas the right-hand side of the equation contains no high-frequency components. Thus, we ignore the first term

via the rotating-wave approximation. Then all that remains is to choose $\omega_{\text{fast},i}(t)$ to cancel out the oscillations of the exponential term, producing a slow bias of the form $r(t)$. One way to obtain this would be to take

$$\omega_{\text{fast},i}(t) = 2r(t) \cos \left[2(\omega_{\text{dc},i} + \delta_i)t + 2 \int_0^t \left(\omega_{\text{slow},i}(t') - \frac{g_i^2}{\omega_{\text{bus}} - \omega_{\text{dc},i} - \omega_{\text{slow},i}(t')} \right) dt' \right], \quad (56)$$

using $\omega_{\text{dc},i}$ and $\omega_{\text{slow},i}(t)$ from (53). There is also a remaining oscillation of the phase at frequencies around $4\omega_{\text{dc}}$, but those can also be neglected with a rotating wave approximation.

Equations (53) and (56), together with the Floquet design parameters $F_i^{(k)}$, constitute a proposed solution to the bus-coupled JPO control problem for achieving dynamical coupling by designing the instantaneous frequency in the JPOs.

1. Solving for the JPO control signals

Here we summarize the procedure for obtaining the time-dependent instantaneous frequencies $\omega_i(t)$, as depicted, for example, in Sup. Figure 1 of the main manuscript. The JPO control problem described by (53) and (56) can be explicitly solved as follows.

We assume that for a given problem size N , problem matrix C , bus frequency ω_{bus} , maximum detuning Δ_{max} , and maximum bus-oscillator coupling g_{max} , we have computed the values for $r_{\text{max}} = \tilde{r}_{\text{max}}\chi$, $\delta_i = \tilde{\delta}_i\chi$, $\Lambda = \tilde{\Lambda}\chi$, $g_i = \tilde{g}_i\chi$, $\Delta_{\text{nom},i} = \tilde{\Delta}_{\text{nom},i}\chi$, and $F_j^{(k)}$ using the prescription described in Sup. Note 5 A. This makes use of various approximations and analytics that are described in later sections below.

1. The first step is to choose

$$\omega_0 = \omega_{\text{bus}} - \Delta_{\text{nom},1} \quad (57)$$

and solve (53a) for $\omega_{\text{dc},i}$ for each i using the quadratic formula.

2. Then, with the solution to $\omega_{\text{dc},i}$, we can solve (53b) for $\omega_{\text{slow},i}(t)$, again using the quadratic formula, for each i and every t .
3. Third, we can insert the above solutions for $\omega_{\text{dc},i}$ and $\omega_{\text{slow},i}$ into (56), and numerically evaluate the integral to find $\omega_{\text{fast},i}(t)$ for each i and every t .
4. Finally, the control signal for the i th JPO $\omega_i(t)$ is, using the two-band approximation (48),

$$\omega_i(t) = \omega_{\text{dc},i} + \omega_{\text{slow},i}(t) + \omega_{\text{fast},i}(t). \quad (58)$$

In Sup. Figure 1, we use $N = 5$, $\omega_{\text{bus}}/2\pi = 10$ GHz, $\Delta_{\text{max}}/2\pi = 5$ GHz, and $g_{\text{max}}/2\pi = 20$ MHz. We choose an SK7 problem instance C that results in prominent sidebands (for the purposes of illustration), which happens to be

$$C = \frac{1}{7} \begin{pmatrix} 0 & 2 & 1 & -4 & 4 \\ 2 & 0 & -5 & 7 & -6 \\ 1 & -5 & 0 & 5 & 3 \\ -4 & 7 & 5 & 0 & -5 \\ 4 & -6 & 3 & -5 & 0 \end{pmatrix}. \quad (59)$$

To visualize the signals in Sup. Figure 1B (right), we compute the power spectral density of

$$x_i(t) := \cos \left[\int_0^t dt' \omega_i(t') \right], \quad (60)$$

which can be interpreted as the in-phase component of an oscillator with instantaneous frequency $\omega_i(t)$.

Finally, we note that for the value of η prescribed by (24), the modulation depth is strong enough to produce asymmetry in the sidebands, which often occurs for phase-modulated signals; for illustrative purposes, however, we reduced this asymmetry by arbitrarily using $\eta = 0.2/N$ instead; all other variables follow the usual procedure.

Supplementary Note 4. Native coupling matrix

Here, we discuss the structure of the native coupling matrix J_{ij} . As alluded to in the previous section, the exact form of $J_{ij}(t)$ is unknown until we have chosen the instantaneous frequencies $\omega_i(t)$; in our dynamical-coupling approach, these are parameterized by the Floquet design parameters $F_j^{(k)}$, which in turn require knowledge of J_{ij} to obtain. One way we can get around this set of coupled dependencies is to constrain ourselves to an approximate form for $J_{ij}(t)$ that is *independent* of $F_i^{(k)}$.

We first address the time dependence of $J_{ij}(t)$. Using (54b), $\omega_{\text{slow } i}(t)$ can be heuristically bounded by

$$|\omega_{\text{slow } i}| \lesssim |\omega_{\text{slow } 1}| \lesssim \max_{t \in (0, 2\pi/\Lambda)} \left| \sum_{k=1}^{N-1} f_1^{(k)} \cos(k\Lambda t) k\Lambda \right| \quad (61)$$

$$\lesssim \delta\omega_{\text{bound } 1} := \Lambda \sum_{k=1}^{N-1} k |f_1^{(k)}|, \quad (62)$$

where $f_1^{(k)}$ is the first-order small-modulation solution (10) to the design problem *with J_{ij} assumed uniform and time-independent*, as was done in Sup. Note 1 B for example. This bound is a good heuristic provided that the native couplings are designed to be close to uniform, a condition that we will discuss shortly. Since $f_1^{(k)}$ depends only on the elements C_{ij} (as indicated by (10)), $\delta\omega_{\text{bound } 1}$ can be readily computed. In fact, if we assume (18) as an approximate form for $f_1^{(k)}$ on average over random problem instances, then

$$\delta\omega_{\text{bound } 1} = \begin{cases} \eta\Lambda(N-1)^{3/2} & \text{for SK7} \\ \frac{1}{2}\eta\Lambda(N-1)^2 & \text{for Dense MAX-CUT} \\ 3\eta\Lambda(N-1)^2N^{-1} & \text{for Cubic MAX-CUT} \end{cases} \quad (63)$$

In any case, regardless of how we compute $\delta\omega_{\text{bound } 1}$, we see that if we impose that $\delta\omega_{\text{bound } 1} \ll \omega_{\text{bus}} - \omega_{\text{dc } 1}$, then we can neglect the effects of (54b), upon which (54a) implies that the oscillator detunings are nominally (i.e., approximately) equally spaced and given by

$$\Delta_{\text{nom } i} := \Delta_{\text{nom } 1} + (i-1)\Lambda, \quad (64)$$

where $\Delta_{\text{nom } 1} := \omega_{\text{bus}} - \omega_{\text{dc } 1} = \omega_{\text{bus}} - \omega_0$ is the nominal detuning of the first oscillator. Under this approximation, the native coupling matrix can be approximated using the nominal detunings via

$$J_{ij} \approx \frac{g_i g_j}{2} \left(\frac{1}{\Delta_{\text{nom } i}} + \frac{1}{\Delta_{\text{nom } j}} \right), \quad (65)$$

which is a time-independent expression. Thus, so long as J_{ij} is approximately uniform, we can ensure time-independence by choosing $\delta\omega_{\text{bound } 1}/\Delta_{\text{nom } 1} \ll 1$.

We now turn to the question of how to ensure the matrix J_{ij} is uniform, in order to justify (61). The uniformity is primarily dictated by the distribution of the bus-resonator couplings g_i . Looking at the form of (65), one reasonable choice of the resonator-bus couplings is

$$g_i = \sqrt{\Delta_{\text{nom } i} \lambda_J}. \quad (66)$$

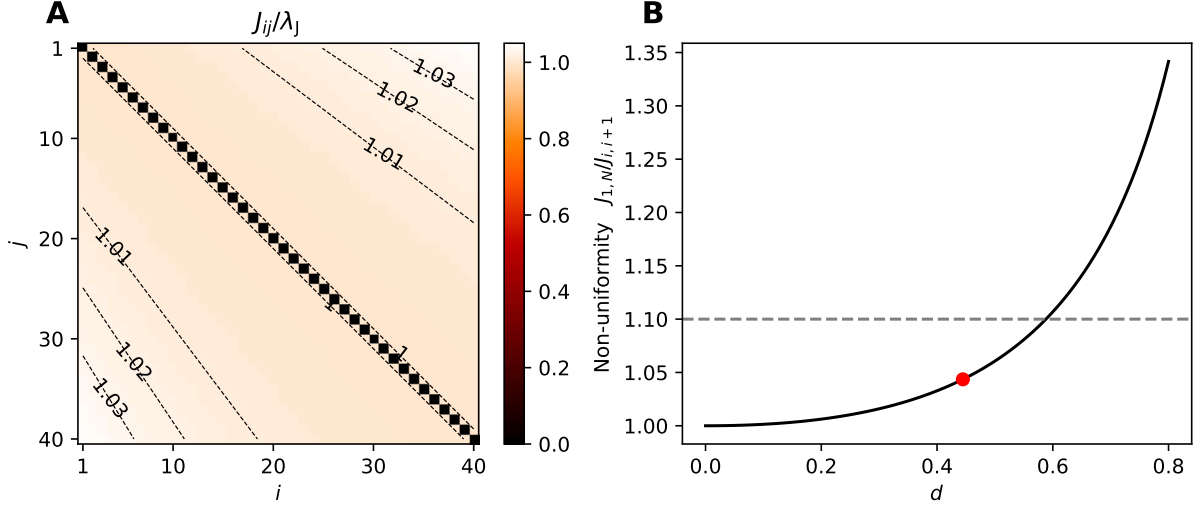
where λ_J is a scalar chosen to represent the approximate value of the elements in the first diagonal of J_{ij} . (If we succeed in obtaining a small non-uniformity, λ_J is also the approximate value of every nondiagonal entry of J .) We show in Sup. Figure 8A the approximation to the native coupling matrix (65) using the choice of resonator-bus coupling (66). We observe that the overall matrix is close in magnitude to λ_J , but there is some degree of non-uniformity.

A good measure for the non-uniformity under this choice of couplings is to look at the ratio between the most off-diagonal element J_{1N} and λ_J , which can be expressed in terms of the nominal detunings $\Delta_{\text{nom } i}$ by

$$\frac{J_{1N}}{\lambda_J} \approx \frac{(\Delta_{\text{nom } 1} + \Delta_{\text{nom } N})/2}{\sqrt{\Delta_{\text{nom } 1} \Delta_{\text{nom } N}}} = \frac{2-d}{2\sqrt{1-d}} \quad (67)$$

where

$$d := \frac{\Delta_{\text{nom } N} - \Delta_{\text{nom } 1}}{\Delta_{\text{nom } N}} \quad (68)$$



Supplementary Figure 8. Structure of native coupling matrix. (A) The approximation to the native coupling matrix J_{ij} given by (65) for $N = 40$ and with nominal detunings set via (70) for a filling ratio of $d = 0.45$. (B) Measure of non-uniformity in the native coupling matrix (67) as a function of the filling ratio d . The red dot represents the point $d = 0.45$ used in (A), while the dashed line indicates a nominal non-uniformity of 10 %, corresponding to $d \approx 0.6$.

is the filling ratio, a dimensionless quantity that characterizes how much of the bandwidth between ω_{bus} and $\Delta_{\text{nom } N}$ is utilized by the oscillators. In Sup. Figure 8B, we plot (67) and see that the non-uniformity scales with d . More concretely, if we allow for a maximum non-uniformity of, say, 10 %, the filling ratio should be at most 0.6. Thus, so long as J_{ij} is approximately time-independent, we can ensure uniformity by choosing d sufficiently small.

Having considered the above, we want to work in a regime in which J_{ij} is both approximately time-independent and uniform. Together with the Schrieffer-Wolff requirements, these considerations can be summarized by a set of constraints involving d . We introduce parameters

1. $\epsilon_{\text{sw}} := g_1/\Delta_{\text{nom } 1}$, which specifies the tolerance for the Schrieffer-Wolff condition $g_1/(\omega_{\text{bus}} - \omega_i) \ll 1$ in the uniform, time-independent approximation;
2. $\epsilon_{\text{ti}} := \delta\omega_{\text{bound } 1}/\Delta_{\text{nom } 1}$, which specifies the tolerance for time-independence in the uniform approximation; and
3. $\epsilon_{\text{uni}} := (2-d)/(2\sqrt{1-d}) - 1$, which specifies the tolerance for uniformity in the time-independent approximation.

If we recast all these parameters in terms of d , we see that in order to satisfy all three constraints, d should satisfy

$$d = \min \left[1 - \left(1 + \epsilon_{\text{sw}}^2 \frac{(N-1)\Lambda}{\lambda_J} \right)^{-1}, 1 - \left(1 + \epsilon_{\text{ti}} \frac{(N-1)\Lambda}{\delta\omega_{\text{bound } 1}} \right)^{-1}, 1 - \left(1 + \sqrt{(1 + \epsilon_{\text{uni}})^2 - 1} + \epsilon_{\text{uni}} \right)^{-2} \right]. \quad (69)$$

In the above, we only technically need d to be bounded by the right-hand side (i.e., a smaller d still satisfies all three constraints). However, that would result in a smaller nonlinear rate χ (as will be discussed in Sup. Note 5 C 3), which in turn implies more stringent requirements on the photon decay rate. Thus, we choose d to saturate its bound above.

Having chosen d , the nominal detunings are then given by

$$\Delta_{\text{nom } i} = \frac{1-d}{d}(N-1)\Lambda + (i-1)\Lambda. \quad (70)$$

Supplementary Note 5. Scaling Considerations for dynamical coupling

Having discussed now our principles for achieving dynamical coupling as well as the key features of the hardware model we would like to demonstrate our principles on, we address in this section how the various parameters, approximations, and constraints come together to specify a concrete system and dynamical-control scheme in a bus-coupled JPO network. First, we summarize our approach in a prescription for choosing dimensionless parameters given any problem instance. Then we analyze how the nonlinear timescale χ should be chosen, which sets the dimensionful realization of a JPO network for consideration. Finally, we analyze how our approach scales with the problem size, particularly in terms of requirements on the photon decay rate.

A. Summary and prescription of system parameters

Up to the nonlinear timescale χ , we are able to supply values for the dimensionless system parameters in order to satisfy all the conditions needed to apply our dynamical-coupling scheme to a bus-coupled system of JPOs. Given the desired target problem couplings C_{ij} with problem size N , we use the following prescription throughout this work:

1. Compute the problem-strength parameter $\tilde{\lambda}_C := \lambda_C/\chi$ via (32) in Sup. Note 2 A:

$$\tilde{\lambda}_C = \begin{cases} 4/N & \text{for SK7} \\ 8/N & \text{for Dense MAX-CUT} \\ 2 & \text{for Cubic MAX-CUT} \end{cases}. \quad (71)$$

This choice ensures the coupling term in the target Hamiltonian scales appropriately with problem size.

2. Set the maximum pump rate $\tilde{r}_{\max} := r_{\max}/\chi$ to its optimal value $\tilde{r}_{\max} = 5$, as discussed in Sup. Note 2 A. This choice is empirically found to work well for the target quantum annealer.
3. Compute the oscillator frame-detunings $\tilde{\delta}_i := \delta_i/\chi$ via (29) in Sup. Note 2 A:

$$\tilde{\delta}_i = \tilde{\lambda}_C \sum_j |C_{ij}|. \quad (72)$$

This choice ensures that the target Hamiltonian operates correctly as a continuous-variable quantum annealer.

4. Compute the dynamical-coupling parameter η via (24) in Sup. Note 1 E:

$$\eta = \begin{cases} 0.8 \cdot N^{-1} & \text{for SK7} \\ 1.2 \cdot N^{-1}(1 + \log N)^{-1} & \text{for Dense MAX-CUT} \\ 0.12 \cdot (1 + \log N)^{-1} & \text{for Cubic MAX-CUT} \end{cases}. \quad (73)$$

This choice ensures that the Floquet design equation can be numerically solved with acceptable error $\mathcal{E}_C \lesssim 3\%$.

5. Compute the dimensionless native-coupling strength parameter via $\tilde{\lambda}_J = \tilde{\lambda}_C/\eta$. This parameter controls the nominal strength of the native couplings J_{ij} and follows from the definition of η .
6. Compute the dimensionless Floquet frequency $\tilde{\Lambda} := \Lambda/\chi$ via (25) in Sup. Note 1 F (specialized here for the JPO system):

$$\tilde{\Lambda} = A \times \max(1, \tilde{\delta}_i, \tilde{r}_{\max}, \tilde{\lambda}_J), \quad (74)$$

where $A \gtrsim 100$; this value of A is empirically found to ensure that the Floquet approximation holds.

7. Compute the dimensionless approximate bound on the slow component of the instantaneous frequency $\tilde{\delta\omega}_{\text{bound1}} := \delta\omega_{\text{bound1}}/\chi$ via (63) in Sup. Note 4:

$$\tilde{\delta\omega}_{\text{bound1}} = \begin{cases} \eta \cdot \tilde{\Lambda} \cdot (N-1)^{3/2} & \text{for SK7} \\ \frac{\eta}{2} \cdot \tilde{\Lambda} \cdot (N-1)^2 & \text{for Dense MAX-CUT} \\ 3\eta \cdot \tilde{\Lambda} \cdot (N-1)^2 N^{-1} & \text{for Cubic MAX-CUT} \end{cases}. \quad (75)$$

This quantity estimates the time dependence of the native coupling using a first-order, uniform approximation for the modulations.

8. Compute the filling ratio d via (69) in Sup. Note 4:

$$d = \min \left[1 - \left(1 + \epsilon_{\text{sw}}^2 \frac{(N-1)\tilde{\Lambda}}{\tilde{\lambda}_{\text{J}}} \right)^{-1}, 1 - \left(1 + \epsilon_{\text{ti}} \frac{(N-1)\tilde{\Lambda}}{\tilde{\delta\omega}_{\text{bound}1}} \right)^{-1}, 1 - \left(1 + \sqrt{(1 + \epsilon_{\text{uni}})^2 - 1} + \epsilon_{\text{uni}} \right)^{-2} \right]. \quad (76)$$

where ϵ_{sw} , ϵ_{ti} , and ϵ_{uni} must be small. (In this work, we use $\epsilon_{\text{sw}} = \epsilon_{\text{ti}} = \epsilon_{\text{uni}} = 1/10$.) This parameter controls the relative magnitudes of the bus-oscillator detunings in order to ensure that the resulting native couplings can be approximated as both uniform and time-independent (in addition to obeying the Schrieffer-Wolff condition).

9. Compute the dimensionless nominal detunings $\tilde{\Delta}_{\text{nom}i} := \Delta_{\text{nom}i}/\chi$ via (70) in Sup. Note 4:

$$\tilde{\Delta}_{\text{nom}i} = \tilde{\Lambda} \left[\frac{1-d}{d} (N-1) + i-1 \right]. \quad (77)$$

These nominal detunings approximate the bus-oscillator detunings as evenly spaced after ignoring small shifts due to coupling non-uniformity and time-dependence.

10. Compute the dimensionless bus-resonator coupling strengths $\tilde{g}_i = g_i/\chi$ via (66) in Sup. Note 4:

$$\tilde{g}_i = \sqrt{\tilde{\Delta}_{\text{nom}i} \tilde{\lambda}_{\text{J}}}. \quad (78)$$

This form for the coupling produces a set of native couplings which approximately have the uniform strength λ_{J} , provided that the $\tilde{\Delta}_{\text{nom}i}$ and d have been chosen accordingly.

11. Approximate the dimensionless native couplings $\tilde{J}_{ij}(t) := J_{ij}(t)/\chi$ by (65) in Sup. Note 4:

$$\tilde{J}_{ij} \approx \frac{\tilde{\lambda}_{\text{J}}}{2} \frac{\tilde{\Delta}_{\text{nom}i} + \tilde{\Delta}_{\text{nom}j}}{\sqrt{\tilde{\Delta}_{\text{nom}i} \tilde{\Delta}_{\text{nom}j}}}. \quad (79)$$

Although the true native couplings $\tilde{J}_{ij}(t)$ are functions of $\Delta_i(t)$ and are thus functions of time, we approximate them by this time-independent expression, which is valid provided that $\tilde{\Delta}_{\text{nom}i}$ are chosen according to the steps above; in this latter case, this approximation also has the property that $\tilde{J}_{ij} \approx \lambda_{\text{J}}$.

12. Solve for the Floquet design parameters $F_i^{(k)}$ with the procedure(s) outlined in Sup. Note 1. To do so, make the formal replacement $\lambda_{\text{C}}/J_{ij} \mapsto \tilde{\lambda}_{\text{C}}/\tilde{J}_{ij}$ in the Floquet integral (5b), where (79) is used to approximate $\tilde{J}_{ij}$. This latter step is a good approximation provided that d has been chosen accordingly, and it crucially allows us to sidestep the problem of $F_j^{(k)}$ being dependent on $\Delta_i(t)$, which is in turn dependent on $F_j^{(k)}$ and so on. Note that for scalability and sufficiently small error $\mathcal{E}_{\text{C}} \lesssim 3\%$ as discussed in Sup. Note 1 E, the full-order objective function (13) should be optimized when $N < 100$, while for $N > 100$, second-order approach (14) should be used instead.

By following this above prescription for any given problem instance, we arrive at a concrete set of dimensionless parameters (namely, $\tilde{\delta}_i$, $\tilde{r}_{\text{max}}$, $\tilde{g}_i$, $\tilde{J}_{ij}$, and $F_j^{(k)}$, the others being secondary to these) which allow us to construct a bus-mediated, dynamically-coupled JPO quantum annealing Hamiltonian (2), at least up to an overall timescale χ . This prescription is thus sufficient for performing simulations (e.g., as discussed in Sup. Note 6 A) to study the performance of the scheme. Furthermore, after selecting the timescale χ as described below, we also arrive at a concrete hardware realization, complete with control signals (as parameterized by $F_j^{(k)}$)—additionally see Sup. Note 3 C for more details on this latter point.

It is worth emphasizing again that this prescription is characterized by our particular approach to obtaining the Floquet design parameters $F_j^{(k)}$: we solve the Floquet design problem using a nominal, time-independent approximation for the native couplings J_{ij} rather than their “true” values $J_{ij}(t)$. To enable this, the other system parameters (e.g., g_i and so on) have been chosen in very particular ways (e.g., through d and its constrained values) such that this approximation holds well. As such, for the purposes of running simulations, we take the approximation (79) to hold *exactly*, and while the design parameters $F_j^{(k)}$ resulting from this procedure could, in principle, be further refined by iterated time-dependent optimization, we forgo doing so for both our simulations and our analysis of hardware realizations/scaling.

Finally, we note that the above prescription is technically constructed for a *lossless* system, in that the cavity-photon decay rate κ is not considered to affect any of the dimensionless parameters mentioned. This goes hand-in-hand with the fact that the parameter T_{ramp} —which governs how adiabatically the pump is ramped up—also has not been discussed, as it should ideally be set to an arbitrarily high value to ensure adiabaticity. However, when the timescale $\tilde{\kappa} := \kappa/\chi$ is also brought into consideration, we prescribe that one should pick

$$\tilde{T}_{\text{ramp}} := T_{\text{ramp}}/\chi = 2/\tilde{\kappa} \quad (80)$$

to obtain order unity number of photon emission events per oscillator, in accordance with the observations made in Sup. Note 2 A. In addition, when photon loss is encountered, one can also empirically tweak the above-prescribed values $\tilde{r}_{\text{max}}$ and $\tilde{\lambda}_J$ (among others) to further enhance success probability. Although we do not include such hyper-optimization here as part of any general procedure, we note that we do incorporate such an approach in some of our quantum simulations (see Sup. Note 6 A for more details).

B. Nonlinear (Kerr) rate

Up to this point, all our discussions have been implicitly in units of the nonlinear rate χ . Thus, selecting a value for χ produces an instantiation of all the other parameters in a dimensionful hardware system. Provided a fixed photon decay rate, it is clear that we want to maximize the value of χ in order to obtain fast system rates relative to the decay rate. However, we cannot pick the rate arbitrarily high, as it is limited by two experimental constraints:

1. $g_i \leq g_{\text{max}}$, the maximal bus-resonator coupling achievable, and
2. $\Delta_N \leq \Delta_{\text{max}}$, the maximal detuning possible due to bandwidth constraints.

To treat the first constraint, we use our form for the bus-oscillator couplings (66) to write it as

$$\sqrt{\Delta_{\text{nom } i} \lambda_J} \leq g_{\text{max}}. \quad (81)$$

Since $\Delta_{\text{nom } i} \leq \Delta_{\text{nom } N}$ by the construction in (70), the oscillator with the largest detuning (i.e., $i = N$) has the largest bus-resonator coupling, and it is thus the most relevant oscillator for this constraint. We substitute the expression for $\Delta_{\text{nom } N}$ from (77) into the constraint above and express the nonlinear rate in terms of dimensionless variables, namely $\chi = \Lambda/\tilde{\Lambda} = \lambda_J/\tilde{\lambda}_J$. Doing this, we arrive at the constraint

$$\chi \leq g_{\text{max}} \sqrt{\frac{d}{\tilde{\Lambda} \tilde{\lambda}_J (N-1)}}. \quad (82)$$

To treat the second constraint, we approximate the detunings Δ_i by the nominal (i.e., evenly spaced) detunings $\Delta_{\text{nom } i}$. Then using $\Delta_N \approx \Delta_{\text{nom } N}$ in the second constraint and using again (77), we arrive at

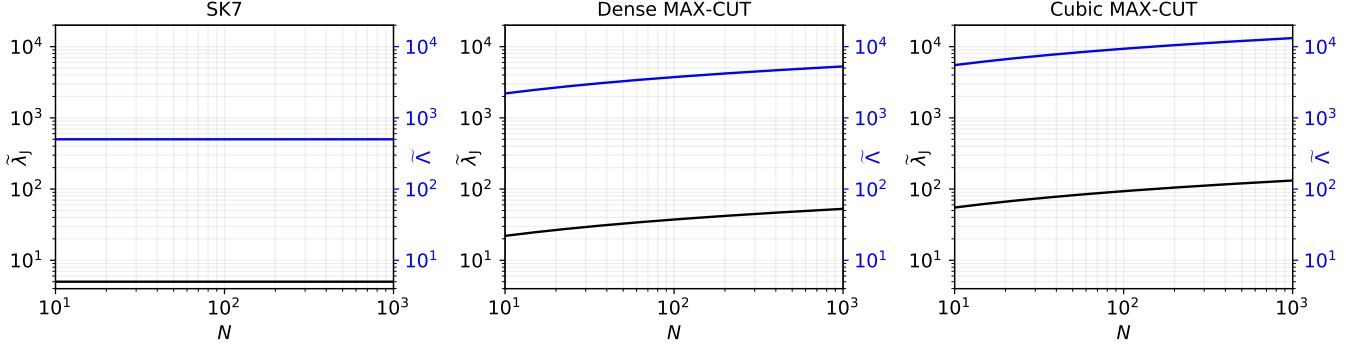
$$\chi \lesssim \Delta_{\text{max}} \frac{d}{(N-1)\tilde{\Lambda}}. \quad (83)$$

Thus, to satisfy both constraints (82) and (83) simultaneously, the nonlinear rate should be given by

$$\chi \leq \chi_{\text{max}} \equiv \min \left(g_{\text{max}} \sqrt{\frac{d}{\tilde{\Lambda} \tilde{\lambda}_J (N-1)}}, \Delta_{\text{max}} \frac{d}{(N-1)\tilde{\Lambda}} \right), \quad (84)$$

where χ_{max} is the maximum allowable nonlinear rate (absolute-scale) before the physical constraints on the bus-resonator couplings and the maximal detuning are violated. As previously mentioned, given a fixed decay rate, it is preferable to set the nonlinear rate to the maximum, i.e., $\chi = \chi_{\text{max}}$. However, this is not technically necessary; for example, if, in a particular hardware implementation, κ is linked to χ in such a way that decreasing χ yields an improvement in χ/κ , such an approach would be compatible with all the constraints we have considered.

As we will see in the following subsection, χ_{max} generally decreases with the problem size N at sufficiently large N . This can be contrasted with the typical scenario for gate-based transmon qubit architectures, where it is desirable to use the maximum nonlinear rate allowed by the hardware (while maintaining acceptable decay rates), more or less independently of N . Instead, in our dynamical-coupling scheme at sufficiently large N , the nonlinear rate χ is forced to be generally much smaller than the hardware-limited maximum, essentially because of the need to fit N oscillators, all distinct in frequency, into a limited bandwidth while still obeying the Floquet approximation, etc. Thus, at least at large N , the key experimental challenge is to obtain decay rates low enough to support such a small nonlinear rate (i.e., at most χ_{max}).



Supplementary Figure 9. Scaling of native coupling strength and Floquet frequency. In black, the dimensionless native coupling strength $\tilde{\lambda}_J$ following (85), and in blue, the dimensionless Floquet frequency $\tilde{\Lambda}$ following (74), both as functions of problem size N . These results follow the prescription outlined in Sup. Note 5 A, wherein $r_{\max} = 5$ and $A = 100$.

C. Scaling of dynamical coupling scheme with problem size

In this subsection, we explore quantitatively the implications of the above results for the scaling of the dynamical coupling scheme with problem size (at least for the JPO networks we consider in this work). To do so, we analyze the scaling of the maximum allowable nonlinear rate $\chi_{\max}$ of the system since this parameter ultimately sets the requirements on the decay rate. In consideration of (84), we first discuss the scaling of the three dimensionless parameters on which $\chi_{\max}$ depends: $\tilde{\lambda}_J$, $\tilde{\Lambda}$, and d .

1. Scaling of native coupling strength and Floquet frequency with problem size

First, we consider the nominal native coupling strength $\tilde{\lambda}_J$. Using the forms we prescribe for $\eta = \tilde{\lambda}_C/\tilde{\lambda}_J$ in (73) and $\tilde{\lambda}_C$ in (71), we get

$$\tilde{\lambda}_J = \begin{cases} 5 & \text{for SK7} \\ (20/3)(1 + \log N) & \text{for Dense MAX-CUT} \\ (50/3)(1 + \log N) & \text{for Cubic MAX-CUT} \end{cases} \quad (85)$$

In particular, because $\tilde{\lambda}_J$ shows up in the denominator of (84), this result implies that SK7 problems have the smallest native coupling strength requirement, while Cubic MAX-CUT problems have the largest.

Next, due to the Floquet timescale requirements, $\tilde{\Lambda} = A \max(1, \tilde{\delta}_i, \tilde{r}_{\max}, \tilde{\lambda}_J)$ by (74). In the case where $\tilde{\lambda}_J$ dominates, $\tilde{\lambda}_J$ and $\tilde{\Lambda}$ are linearly related, in which case $\tilde{\lambda}_J$ and $\tilde{\Lambda}$ will have the same functional form. For our prescription, this is the most relevant situation, since $\tilde{r}_{\max}$ is constant with N in our prescription, so unless $\tilde{\Lambda}$ is a constant, it will be limited by $\tilde{\lambda}_J$ at sufficiently large N .

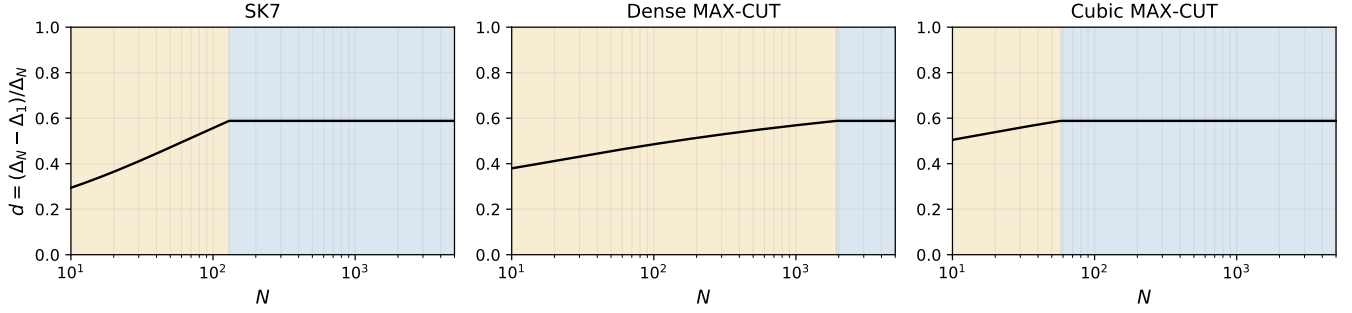
In Sup. Figure 9, we show the values of $\tilde{\lambda}_J$ following (85) and $\tilde{\Lambda}$ following (74) as a function of problem size. From the figure, it turns out $\tilde{\lambda}_J \geq 5$ in all cases, and since $\tilde{r}_{\max} = 5$ in our prescription (see also Sup. Note 2), $\tilde{\lambda}_J$ limits the value of $\tilde{\Lambda}$ in all cases. (Note also that $\tilde{\delta}_i$ does not limit $\tilde{\Lambda}$ provided we follow (71) and (72).)

2. Scaling of filling ratio with problem size

In (84), the value of χ increases for larger filling ratio d , which implies that it is important for d to be large to maximize the performance of the dynamical-coupling scheme. In Sup. Note 4, we showed that the filling ratio is determined by three different constraints, which is summarized by (76) in our prescription. We can rewrite this as

$$d = \min \left[\frac{D_{\text{sw}}}{1 + D_{\text{sw}}}, \frac{D_{\text{ti}}}{1 + D_{\text{ti}}}, d_{\text{uni}} \right], \quad \text{where} \quad (86)$$

$$D_{\text{sw}} := \epsilon_{\text{sw}}^2 (N - 1) \tilde{\Lambda} / \tilde{\lambda}_J, \quad D_{\text{ti}} := \epsilon_{\text{ti}} (N - 1) \tilde{\Lambda} / \tilde{\delta} \omega_{\text{bound}1}, \quad d_{\text{uni}} := 1 - \left(1 + \sqrt{(1 + \epsilon_{\text{uni}})^2 - 1 + \epsilon_{\text{uni}}} \right)^{-2}.$$



Supplementary Figure 10. Scaling of filling ratio. The filling ratio d as given by (86) as a function of N , with D_{sw} given by (87) and D_{ti} given by (88). The gold and blue regions denote where d is constrained by the time-independence and uniformity requirements, respectively. These results follow the prescription outlined in Sup. Note 5 A, wherein $\epsilon_{\text{sw}} = \epsilon_{\text{ti}} = \epsilon_{\text{uni}} = 0.1$.

In this form, we see that the uniformity constraint (dictated by d_{uni}) does not depend on the problem size N . As discussed in Sup. Note 4, this uniformity constraint fundamentally limits the bandwidth utilization to about 60% (i.e., $d \leq d_{\text{uni}} = 0.6$) for a non-uniformity in the native coupling matrix of 10% (i.e., $\epsilon_{\text{uni}} = 0.1$). On the other hand, both the Schrieffer-Wolff and time-independence conditions vary with N and are both described by a similar functional form, where d increases for larger D_{sw} or D_{ti} . Thus, for sufficiently large D_{sw} and D_{ti} , we have the simple picture that the filling ratio d is limited by the uniformity constraint to give $d = d_{\text{uni}}$. To understand when this is the case, we study D_{sw} and D_{ti} below.

The simpler of the two is the Schrieffer-Wolff condition. Assuming, as in the previous subsection, that $\tilde{\Lambda} = A\tilde{\lambda}_J$, we can first rewrite

$$D_{\text{sw}} = 2\epsilon_{\text{sw}}^2 A(N-1). \quad (87)$$

Thus, D_{sw} is linearly proportional to the problem size and does not depend on the problem class. For sufficiently large N , this condition ceases to limit d .

Turning to the time-independence constraint, we can substitute the form for $\tilde{\delta\omega}_{\text{bound}}$ from (75) in our prescription into D_{ti} , upon which we see that D_{ti} depends on N and η . Then using the form for η from (73) in our prescription, we arrive at

$$D_{\text{ti}} = \begin{cases} (5/4)\epsilon_{\text{ti}} \cdot N^{1/2} \cdot \sqrt{N/(N-1)} & \text{for SK7} \\ (5/3)\epsilon_{\text{ti}} \cdot (1 + \log N) \cdot N/(N-1) & \text{for Dense MAX-CUT} \\ (25/9)\epsilon_{\text{ti}} \cdot (1 + \log N) \cdot N/(N-1) & \text{for Cubic MAX-CUT} \end{cases} \quad (88)$$

We can see that D_{ti} increases with the problem size in all cases, so the time-independence constraint also ceases to limit d at sufficiently large N . Moreover, for the SK7 problems, D_{ti} increases as $N^{1/2}$ while for the Dense and Cubic MAX-CUT problems, D_{ti} increases more slowly. This implies that for the SK7 problems, d varies more drastically with respect to N than for Dense and Cubic MAX-CUT problems.

In Sup. Figure 10, we show the filling ratio d as a function of problem size following the above analysis. For the parameters we consider, it turns out that d is never limited by the Schrieffer-Wolff condition. Rather, it is dominated by the time-independence condition at small N before being clamped by the uniformity condition at approximately 60%, as argued above. Finally, these results also verify that d depends more weakly on N for the MAX-CUT problems.

3. Scaling of nonlinear rate with problem size

Putting the above results together via (84) produces the maximum allowable nonlinear rate χ_{max} for each desired problem size N . Since the effects of $\tilde{\lambda}_J$ and $\tilde{\Lambda}$ are to decrease χ_{max} with N , while the effects of d eventually saturates with N , we ultimately expect that χ_{max} decreases with N .

This trend can be quantitatively seen in Sup. Figures 11, 12 and 13. In addition to the general trend of decreasing χ_{max} with N , we also see that the particular scaling behavior of χ_{max} with N changes dramatically when the system switches from being limited by the maximum coupling strength g_{max} (where it is more gentle) to being limited by the maximum available bandwidth Δ_{max} (where it is sharper).

While Sup. Figures 11, 12 and 13 use the exact formula for $\chi_{\max}$ given by (84), we can also qualitatively understand much of the scaling behavior using the arguments already presented. First, based on what we found for the scaling of d , we can make the simplifying assumption that d is constant (this holds at large N , and even when d increases at small N , it is a slow function). Then we can insert the expressions for $\tilde{\lambda}_J$ from (85) along with $\tilde{\Lambda} \propto \tilde{\lambda}_J$ as argued above. In doing so, we find that when $\chi_{\max}$ is constrained by $g_{\max}$, we have (with $N - 1 \approx N$)

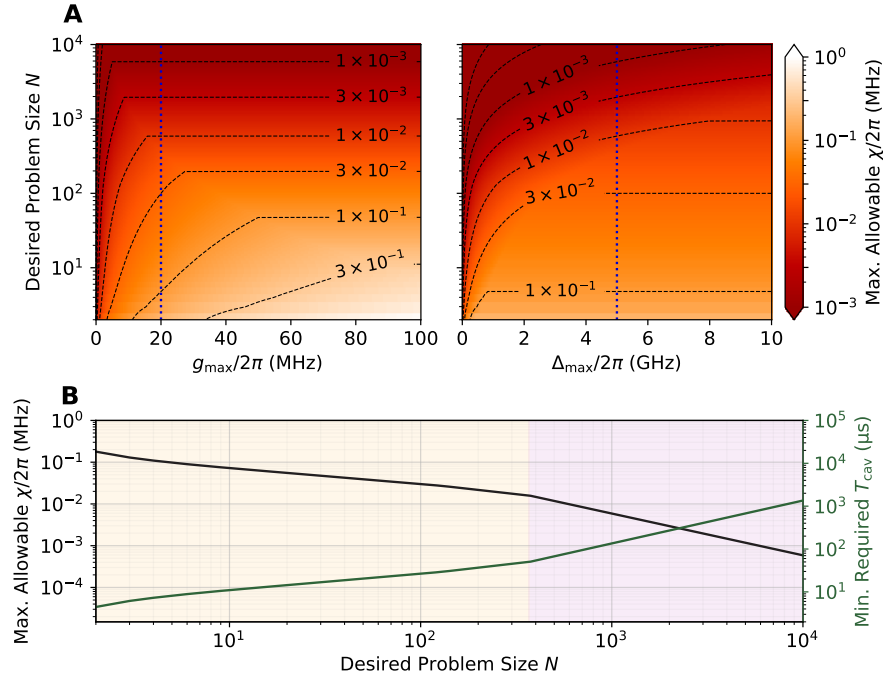
$$\chi_{\max} \propto \begin{cases} N^{-1/2} & \text{for SK7} \\ N^{-1/2} \cdot (1 + \log N)^{-1} & \text{for Dense MAX-CUT} \\ N^{-1/2} \cdot (1 + \log N)^{-1} & \text{for Cubic MAX-CUT} \end{cases} \quad (89)$$

On the other hand, when the system is constrained by $\Delta_{\max}$, we can apply the same simplifying assumptions to get

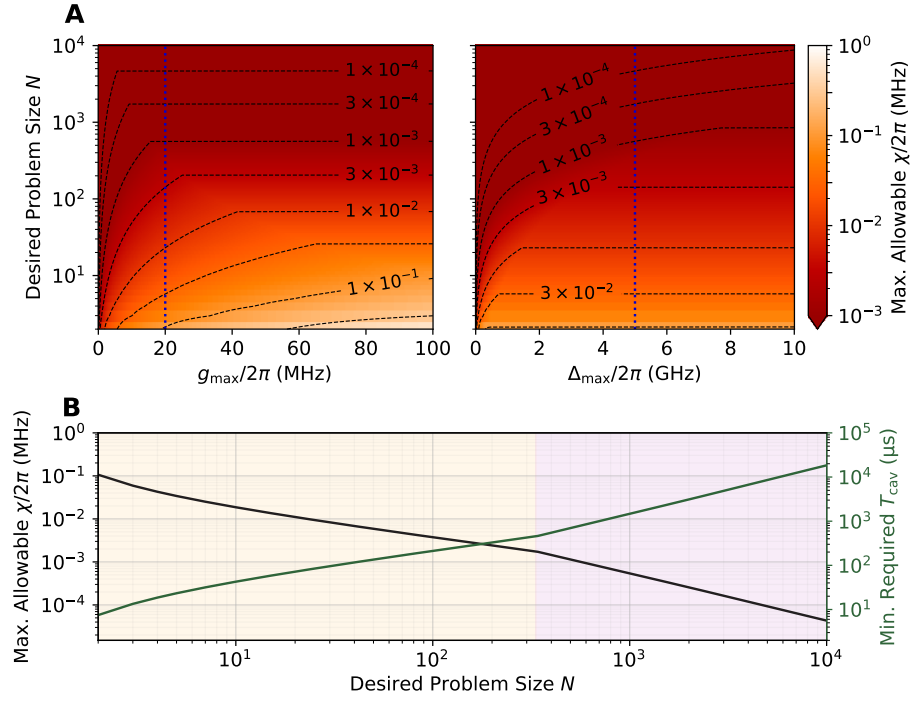
$$\chi_{\max} \propto \begin{cases} N^{-1} & \text{for SK7} \\ N^{-1} \cdot (1 + \log N)^{-1} & \text{for Dense MAX-CUT} \\ N^{-1} \cdot (1 + \log N)^{-1} & \text{for Cubic MAX-CUT} \end{cases} \quad (90)$$

Thus, at large N , the bandwidth constraint due to $\Delta_{\max}$ is the one that limits the scaling of $\chi_{\max}$, while the milder coupling constraint due to $g_{\max}$ is potentially relevant for intermediate values of N . The latter result is especially interesting for near-term demonstrations of dynamical coupling as (89) implies that the scaling requirements for $\chi_{\max}$ (and thus ultimately of the decay rate) are less stringent than the $1/N$ scaling that one might naively expect, and which indeed dominates at large N .

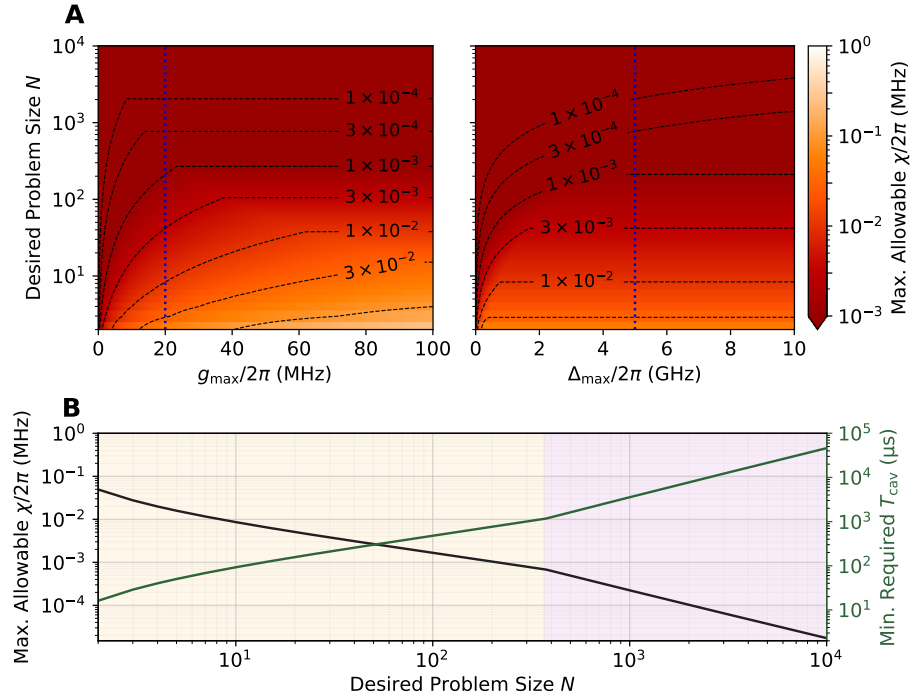
Finally, we note that both (89) and (90) show that the scaling requirements are more favorable for the SK7 problem class. This observation is supported quantitatively by Sup. Figures 11, 12 and 13. This advantage stems from the much smaller Floquet timescale requirement for $\tilde{\Lambda}$ for SK7 problems, as discussed in Sup. Note 1 F.



Supplementary Figure 11. Scaling requirements for SK7 problems. (Reproduced from Sup. Figure 5 of main manuscript) (A) The maximum allowable nonlinear rate $\chi_{\max}$ according to (84), as a function of the desired problem size N and the maximum achievable bus-oscillator coupling rate $g_{\max}$ (left, with $\Delta_{\max}/2\pi = 5$ GHz) and the maximum achievable bus-oscillator detuning $\Delta_{\max}$ (right, with $g_{\max}/2\pi = 20$ MHz). The black dotted lines are level curves of constant $\chi_{\max}$, while the dashed blue lines indicate the slice of these plots shown in (B) below. (B) On the left axis (black curve), $\chi_{\max}$ as a function of problem size for $\Delta_{\max}/2\pi = 5$ GHz and $g_{\max}/2\pi = 20$ MHz. On the right axis (green curve), the minimum required cavity lifetime T_{cav} , assuming we use $\chi = \chi_{\max}$ and that we need $T_{\text{cav}} = 5/\chi$ for the quantum annealer to perform effectively. The beige shaded region indicates the regime in which $\chi_{\max}$ is limited by $g_{\max}$ while the pink shaded region indicates where it is limited by $\Delta_{\max}$. All parameters of the system (e.g., $\tilde{\Lambda}$, $\tilde{\lambda}_J$ and d) follow the prescription specified in Sup. Note 5 A.



Supplementary Figure 12. Scaling requirements for Dense MAX-CUT problems. See caption of Sup. Figure 11.



Supplementary Figure 13. Scaling requirements for Cubic MAX-CUT problems. See caption of Sup. Figure 11.

Supplementary Note 6. Models and Methodologies for Numerical Simulations

In this section, we specify and review the mathematical framework we use to model the quantum states and dynamics of the KPO system, under both static coupling and dynamical coupling. We describe the quantum simulations we perform and the procedure we use to interpret the resulting state trajectories as success probabilities. We also describe the classical EOMs we use to obtain intuition about how the observations made in the quantum simulations scale for larger problem sizes.

A. Quantum Simulations

We perform quantum simulations in this work to analyze the performance of the dynamical-coupling scheme relative to the statically-coupled system, especially in the presence of dissipation. Our dissipation model is described by the standard Lindblad master equation

$$\dot{\hat{\rho}} = -i[\hat{H}, \hat{\rho}] + \sum_m \left(\hat{L}_m \hat{\rho} \hat{L}_m^\dagger - \frac{1}{2} \hat{L}_m^\dagger \hat{L}_m \hat{\rho} - \frac{1}{2} \hat{\rho} \hat{L}_m^\dagger \hat{L}_m \right) \quad (91)$$

where $\hat{H}$ is the Hamiltonian of the system, and $\hat{L}_m$ are the Lindblad operators that describe the effect of dissipation due to coupling to the environment; the time dependence of all operators have been omitted for brevity. For the statically-coupled scheme, the Hamiltonian is given by (26) while the Lindblad operators are given by (30). For the dynamical-coupling scheme, the Hamiltonian is given by (3) in the main manuscript, while the Lindblad operators, due to being in the rotating frame, are given by $\hat{L}_m = \sqrt{\kappa} \hat{a}_m \exp(-i \int_0^t (\omega_m(t') - g_m^2/\Delta_m(t') + \delta_m) dt')$. However, by direct substitution of this Lindblad operator into (91), we see that the overall phase is always canceled with its complex conjugate; this implies that the overall phase in the Lindblad operator does not have any observable consequences for the internal state of the open quantum system. Thus, without loss of generality, we take the Lindblad operator of the dynamical-coupling scheme to be $\hat{L}_m = \sqrt{\kappa} \hat{a}_m$, as for the statically-coupled system.

Directly simulating (91) is difficult due to the large Hilbert-space dimension of the quantum state. To circumvent the memory requirement of storing a full density matrix, we utilize the Monte-Carlo wavefunction method (i.e., quantum jump method) to solve (91) via stochastic sampling [9]. Given a finite number of trajectories N_{traj} , this method approximates the density matrix $\hat{\rho}(t)$ using the ensemble mean of the trajectories $|\psi^{(\ell)}(t)\rangle$ via

$$\hat{\rho}(t) \approx \frac{1}{N_{\text{traj}}} \sum_{\ell=1}^{N_{\text{traj}}} |\psi^{(\ell)}(t)\rangle \langle \psi^{(\ell)}(t)|. \quad (92)$$

It should be noted that for the idealized lossless case of $\kappa = 0$, the number of trajectories should be taken to be $N_{\text{traj}} = 1$ as there is no stochasticity due to the lack of dissipation to the environment; this is equivalent to solving the Schrödinger equation.

Given a spin configuration $\sigma = (\sigma_1, \dots, \sigma_N)$ where $\sigma_i = \pm 1$, we can define, for the ℓ th trajectory, the probability of the quantum state $|\psi^{(\ell)}(t)\rangle$ being in the specific spin configuration σ as [1]

$$P_{\sigma}^{(\ell)}(t) := \prod_{i=1}^N \left(\int_{\text{sgn } x_i = \sigma_i} dx_i \right) \left| \psi^{(\ell)}(x_1, \dots, x_N, t) \right|^2 \quad (93)$$

where $\psi^{(\ell)}(x_1, x_2, \dots, x_N, t) := \langle x_1, x_2, \dots, x_N | \psi^{(\ell)}(t) \rangle$ is the wavefunction in the position basis. (Note: We use the convention $\hat{x}_i := (\hat{a}_i + \hat{a}_i^\dagger)/2$.) For our numerical approach to computing (93), see Sup. Note 6 A 1 below.

This definition of the success probability is motivated by the fact that the final state of the KPO system is approximately a superposition of coherent states $|\pm\alpha_i\rangle$ for each KPO, and we interpret these coherent states to encode the Ising spins $\sigma_i = \pm 1$. In practice, we obtain the result of the computation by an in-phase homodyne measurement of $\hat{x}$, and the sign of the result on the i th KPO is interpreted as σ_i . Under this definition of the spin configuration probability, the vacuum state at the beginning of the computation has equal probability on all possible spin configurations, and as the computation progresses, the displacements of the KPOs evolve—in a correlated way—to remove the ambiguity, pushing the probability distribution into ideally the ground-state quadrants of phase space.

Using (93), we define the success probability of the ℓ th trajectory as

$$P_{\text{success}}^{(\ell)}(t) := \sum_{\sigma \in \mathcal{S}_{\text{gs}}} P_{\sigma}^{(\ell)}(t) \quad (94)$$

Simulation	$\tilde{\lambda}_C$	$\tilde{T}_{\text{ramp}}$	$\tilde{r}_{\text{max}}$
$N = 2, \kappa/\chi = 0$	2	100	5.0
$N = 4, \kappa/\chi = 0$	1	100	5.0
$N = 4, \kappa/\chi = 2 \times 10^{-2}$	1	25	4.5
$N = 4, \kappa/\chi = 2 \times 10^{-1}$	1	5	4.0

Supplementary Table 1. System Parameters. System parameters associated with the static annealer (see Sup. Note 2 A) used for the quantum simulations in this work, chosen to optimize the success probability. These optimized parameters also carry over to the quantum simulation of the corresponding dynamically-coupled system.

where $\mathcal{S}_{\text{gs}}$ is set of spin configurations that are ground states of the specified Ising problem.

These trajectory-wise definitions for the configuration and success probabilities are distinct from the ensemble-averaged configuration and success probabilities which would be computed from (91). We estimate these ensemble-averaged quantities, P_σ and P_{success} , respectively, by taking the sample mean of the trajectory-wise quantities:

$$P_\sigma(t) \approx \frac{1}{N_{\text{traj}}} \sum_{\ell=1}^{N_{\text{traj}}} P_\sigma^{(\ell)}(t), \quad (95)$$

$$P_{\text{success}}(t) \approx \frac{1}{N_{\text{traj}}} \sum_{\ell=1}^{N_{\text{traj}}} P_{\text{success}}^{(\ell)}(t). \quad (96)$$

Of course, since N_{traj} is finite, there is statistical sampling uncertainty on this estimation, which we henceforth take to be the standard error of the mean. We show the standard error of the mean as a colored fill around the mean on Sup. Figures 19–22, 25–28, and 31–34.

Our quantum simulation procedure is as follows. Given a problem matrix C_{ij} , we chose the parameters $\tilde{r}_{\text{max}}$, $\tilde{T}_{\text{ramp}}$, $\tilde{\lambda}_C$ to maximize the success probabilities of the static annealer, summarized in Sup. Table 1 for the simulations in this work. Then we use Steps 3–12 of the prescription outlined in Sup. Note 5 A to calculate the system parameters $\tilde{\delta}_i$, $\tilde{J}$, and $\tilde{\Lambda}$ and $F_j^{(k)}$. Since we take the unit of time to be $1/\chi$, the simulations are independent of dimensionful constraints associated with particular physical realizations (i.e., due to g_{max} and Δ_{max}). Having obtained these dimensionless system parameters, we simulate both the statically- and dynamically-coupled systems using the Monte-Carlo wavefunction method with the QuTiP library (version 4.3.1) in Python [15]. To reliably compare the stochastic results from the statically- and dynamically-coupled systems, we use the same noise process (i.e., random number generator) to ensure that both systems physically experience the same instantiation of quantum noise.

As is the standard approach, we run our simulations with various truncation settings for the Fock basis of each mode, and we choose a setting which does not result in significant changes in the results upon increase. (Specifically, for all the quantum simulations in this work, the truncation was set to 12 basis states per mode, corresponding to 12^4 -dimensional states for $N = 4$ simulations.) As is also standard, we decrease the time step taken in our solver until no significant changes in the result occur upon further decreasing the step size.

This approach was taken to generate Sup. Figure 2 in the main manuscript. For the simulations with no cavity photon loss ($\kappa = 0$), $N_{\text{traj}} = 1$; while for the low-loss ($\kappa = 0.02\chi$) and high-loss ($\kappa = 0.2\chi$) cases, $N_{\text{traj}} = 10$. For the 2-mode wavefunctions shown in Sup. Figure 2A, we use the approach described below in Sup. Note 6 A 1. The success probability used in the histogram is defined as the success probability at the end of the computation $P_{\text{success}} := P_{\text{success}}(T)$. Note that even though the histogram in Sup. Figure 2D does not show the uncertainty of the success probabilities, these uncertainties can be obtained by examining the trajectory plots of each problem instances in Sup. Figures 19–22.

1. Continuous-variable representations of state

One simple way to visualize the correlations among the oscillators is to look at the quantum state in a continuous-variable basis. For example, in position space, the population (i.e., diagonal) elements of the density operator $\hat{\rho}$ can be represented as a function over $x \in \mathbb{R}^N$ as $p_x(x) := \langle x | \hat{\rho} | x \rangle$. (For pure states $\hat{\rho} = |\psi\rangle\langle\psi|$, $p_x(x) = |\langle x | \psi \rangle|^2$.) Of course, to obtain non-classical information about the state, we also need to look at a conjugate variable; for example, we can also consider the momentum-basis population $p_p(p)$, computed in the same way but with $|x\rangle \mapsto |p\rangle$.

In practice, since $\hat{\rho}$ is represented (e.g., in QuTiP) in the Fock basis, a basis transformation is needed to compute such matrix elements. The state $|x\rangle$ can be obtained in the Fock basis via a standard basis transformation:

$$|x\rangle = \sum_{n \in \mathbb{N}^N} \langle n|x\rangle |n\rangle = \bigotimes_{i=1}^N \sum_{n_i=0}^{\infty} \langle n_i|x_i\rangle |n_i\rangle, \quad (97)$$

which can be readily computed since $\langle x_i|n_i\rangle = \phi_{n_i}^{(x)}(x_i)$, where for $x \in \mathbb{R}$,

$$\phi_n^{(x)}(x) := \frac{(-1)^n}{(\sqrt{\pi}2^n n!)^{1/2}} \exp(x^2/2) \frac{d^n}{dx^n} \exp(-x^2) \quad (98)$$

is the position-basis wavefunction of the n th eigenstate of the quantum harmonic oscillator (i.e., the n th Fock state). Similarly, if we are interested in $\langle p|n\rangle$, we can perform the same calculation using the Fock states in the momentum basis instead, using $\phi_n^{(p)}(p) := (-i)^n \phi_n^{(x)}(p)$.

Explicitly, one can in practice (e.g., in QuTiP) compute $p_x(x_1, \dots, x_N)$ by evaluating the “matrix element” of the “density matrix” $P_{n'n} := \langle n'_1, \dots, n'_N | \hat{\rho} | n_1, \dots, n_N \rangle$ against the “state vector” $c_n(x) := \langle n_1, \dots, n_N | x_1, \dots, x_N \rangle$. Furthermore, we note that, in practice, the summations (e.g., in (97)) are limited by truncation in the Fock basis, with upper limits M_n such that the support of $\langle n_1, \dots, n_N | \hat{\rho} | n_1, \dots, n_N \rangle$ is negligible for $n_i > M_i$.

2. Evaluation of success probabilities

A similar consideration is needed to evaluate the success probabilities defined in (93) when the quantum state is represented in the Fock basis. To do so, we note that (focusing on the case of $N = 1$ so that $x \in \mathbb{R}$)

$$\int_{x_{\min}}^{x_{\max}} |\psi(x)|^2 dx = \int_{x_{\min}}^{x_{\max}} dx \langle \psi|x\rangle \langle x|\psi\rangle = \langle \hat{\Pi}_{x_{\min}}^{x_{\max}} \rangle_{|\psi\rangle}, \quad (99)$$

where $\hat{\Pi}_{x_{\min}}^{x_{\max}} := \int_{x_{\min}}^{x_{\max}} |x\rangle \langle x| dx$ is the projector onto the set of x -eigenstates between $x_{\min}$ and $x_{\max}$. We need to obtain the Fock representation of this projector. For any two Fock states $|n'\rangle$ and $|n\rangle$, its matrix elements are

$$\langle n' | \hat{\Pi}_{x_{\min}}^{x_{\max}} | n \rangle = \int_{x_{\min}}^{x_{\max}} \langle n'|x\rangle \langle x|n\rangle dx, \quad (100)$$

which can be readily evaluated using (98). From this, we can continue (99) by writing

$$\langle \hat{\Pi}_{x_{\min}}^{x_{\max}} \rangle_{|\psi\rangle} = \sum_{n'=0}^{\infty} \sum_{n=0}^{\infty} \langle \psi|n'\rangle \langle n' | \hat{\Pi}_{x_{\min}}^{x_{\max}} | n \rangle \langle n|\psi\rangle. \quad (101)$$

Explicitly, one can compute this in practice (e.g. in QuTiP) as an “expectation value” of the “operator matrix” $(P_{x_{\min}}^{x_{\max}})_{n'n} := \langle n' | \hat{\Pi}_{x_{\min}}^{x_{\max}} | n \rangle$ on the “state vector” $c_n := \langle n|\psi\rangle$.

For $N > 1$, the system composition rule of quantum mechanics implies that

$$\prod_{i=1}^N \int_{x_{\min i}}^{x_{\max i}} dx_i |\psi(x_1, \dots, x_N)|^2 = \prod_{i=1}^N \sum_{n_i, n'_i=0}^{\infty} \overbrace{\langle n'_1 | \hat{\Pi}_{x_{\max 1}} | n_1 \rangle \cdots \langle n'_N | \hat{\Pi}_{x_{\max N}} | n_N \rangle}^{[P_{x_{\min 1}}^{x_{\max 1}} \otimes \cdots \otimes P_{x_{\min N}}^{x_{\max N}}]_{n'_1, \dots, n'_N, n_1, \dots, n_N}} \overbrace{\langle \psi | n'_1, \dots, n'_N \rangle}^{c_{n'_1, \dots, n'_N}^*} \overbrace{\langle n_1, \dots, n_N | \psi \rangle}^{c_{n_1, \dots, n_N}}, \quad (102)$$

where $\otimes$ indicates the Kronecker product of matrices. Again, as indicated by the braced quantities above the expression, this quantity can be evaluated in practice (e.g., in QuTiP) as an expectation value of an operator matrix on a state vector (both over an N -mode Fock space).

In practice, there are two more issues we need to address: discretization and truncation in position basis. (Truncation in the Fock basis has already been discussed above and simply consists of restricting the upper limit of all the summations.) In order to evaluate the integral in (100), we can discretize the interval $x_{\min} < x < x_{\max}$ into L sections, and we approximate the matrix element by

$$(P_{x_{\min}}^{x_{\max}})_{n'n} \approx \sum_{j=0}^{L-1} \langle n'|x^{(j)}\rangle \langle x^{(j)}|n\rangle \Delta x, \quad (103a)$$

where

$$\Delta x := \frac{x_{\max} - x_{\min}}{L} \quad \text{and} \quad x^{(j)} := x_{\min} + \left(j + \frac{1}{2}\right) \Delta x. \quad (103b)$$

Note that if our Fock basis is truncated to M basis elements, (103a) is simply a multiplication of an $M \times L$ matrix with an $L \times M$ matrix (the latter representing the change of basis from truncated-Fock to discretized-position, and the former vice-versa). Finally, the position basis is also truncated in practice, so to represent the improper integral $\int_{\mathbb{R}^+} |\psi(x)|^2 dx$, we need to pick $x_{\min} = 0$ and $x_{\max}$ sufficiently large as to cover the non-negligible support of $\psi(x > 0)$, while to represent $\int_{\mathbb{R}^-} |\psi(x)|^2 dx$, we pick $x_{\min}$ sufficiently negative as to cover the non-negligible support of $\psi(x < 0)$ and $x_{\max} = 0$.

B. Classical Simulations

We also perform classical simulations to analyze how the performance of the dynamical-coupling scheme scales with respect to problem size N . The idea here is to derive equations of motion (EOMs) with system dimension linear in N , in order to make numerical analysis tractable. We start with the quantum Heisenberg-Langevin EOMs

$$\frac{d\hat{a}_i}{dt} = -i[\hat{a}_i, \hat{H}] + \frac{1}{2} \sum_m \left(\hat{L}_m^\dagger [\hat{a}_i, \hat{L}_m] + [\hat{L}_m^\dagger, \hat{a}_i] \hat{L}_m \right) + \text{Noise terms} \quad (104)$$

describing the quantum system and then perform the formal substitution $\hat{a}_i \rightarrow \alpha_i$, where $\alpha_i \in \mathbb{C}$; the noise terms are then neglected. Although these classicalized EOMs are not a physically correct model for the KPO network in the quantum regime, such an approach has been used in, e.g., Ref. [4] to understand the operational mechanism of such quantum devices, and models of this kind have been shown to be effective heuristic Ising solvers in and of themselves [16]. We utilize the same approach to help inform the scaling of the Floquet requirements at larger problem sizes (of the quantum model).

For the statically-coupled system, this classicalization procedure produces the EOMs

$$\frac{1}{\chi} \frac{d\alpha_i}{dt} = -\frac{\tilde{\kappa}}{2} \alpha_i + i \left(\tilde{\delta}_i + |\alpha_i|^2 \right) \alpha_i - i\tilde{r}\alpha_i^* + i\tilde{\lambda}_C \sum_{i \neq j} C_{ij} \alpha_j. \quad (105)$$

Using the same procedure, the dynamically-coupled system has the following EOMs:

$$\frac{1}{\chi} \frac{d\alpha_i}{dt} = -\frac{\tilde{\kappa}}{2} \alpha_i + i \left(\tilde{\delta}_i + |\alpha_i|^2 \right) \alpha_i - i\tilde{r}\alpha_i^* + i\tilde{\lambda}_C \sum_{j \neq i} J_{ij} \exp[-i\Lambda(j-i)t + i\delta\phi_j(t) - i\delta\phi_i(t)] \alpha_j. \quad (106)$$

For all the classical simulations, we use $\tilde{\kappa} = 0$. From empirical investigations, the value of $\tilde{\kappa}$ does not significantly affect the qualitative distribution of the success probabilities in this classicalized model.

Since $\alpha_i = 0$ is an unstable equilibrium, we follow the approach of Reference [1] and sample random initial-field amplitudes $\alpha_i(t=0)$, according to

$$\alpha_i(t=0) = \sqrt{\frac{n_0}{2}} z_i e^{2i\pi u_i}, \quad (107)$$

where z_i are iid random variables with the standard normal distribution while u_i are iid random variables with the uniform distribution on the interval $[0, 1]$. For all classical simulations, we choose $n_0 = 10^{-2}$.

As with the quantum simulations, we simulate the classical EOMs on a finite number of trajectories N_{traj} . The ℓ th trajectory $\alpha^{(\ell)}(t)$ is an N -dimensional vector at each time point t , and each ℓ corresponds with a random sample from (107) for the initial condition of the trajectory. Note that in contrast to the quantum case, we always need to take $N_{\text{traj}} > 1$ in all cases, even though $\tilde{\kappa} = 0$.

In this approach, we identify the spin configuration of a trajectory as a function of time using the relation

$$\sigma_i^{(\ell)}(t) = \text{sgn}(\text{Re}(\alpha_i^{(\ell)}(t))). \quad (108)$$

The success of a given trajectory ℓ at time t is then given by the indicator variable

$$s^{(\ell)}(t) := \begin{cases} 1 & \text{if } \sigma^{(\ell)}(t) \in \mathcal{S}_{\text{gs}}, \\ 0 & \text{if } \sigma^{(\ell)}(t) \notin \mathcal{S}_{\text{gs}}, \end{cases} \quad (109)$$

where $\mathcal{S}_{\text{gs}}$ is the set of ground-state spin configurations. Then, as in the quantum case, we define the success probability of the ensemble of trajectories to be the sample mean

$$P_{\text{success}}(t) = \frac{1}{N_{\text{traj}}} \sum_{\ell} s^{(\ell)}(t). \quad (110)$$

Again, there is a statistical sampling uncertainty associated with the sample mean, which we take to be the usual standard error on the mean. We show the standard error of the mean as a colored fill around the mean on Sup. Figure 3A.

Our classical simulation procedure is as follows. Given a problem matrix C , we use step 1–12 in the prescription outlined in 5 A to calculate the system parameters $\tilde{r}_{\text{max}}$, $\tilde{\delta}_i$, $\tilde{\lambda}_C$, $\tilde{J}$, $\tilde{T}$ and $\tilde{\Lambda}$ and $F_i^{(k)}$. Since we take the unit of time to be $1/\chi$, the simulations are independent of dimensionful constraints associated with particular physical realizations (i.e., due to g_{max} and Δ_{max}). Finally, for simplicity, we use $\tilde{\kappa} = 0$ for these simulations.

For each trajectory index, both the statically- and dynamically-coupled systems are simulated with the same (random) initial field amplitude. We then simulate both the statically- and dynamically-coupled systems using the ODE solver library DifferentialEquations.jl (version 4.5.0) in Julia [17]. As is the standard approach, we decrease the time step taken in our solver until no significant changes in the results occur upon further decrease of the step size. Note that since the timescale of the dynamics (i.e., $1/\tilde{\Lambda}$) become smaller with increasing N , the time step required for convergence decreases as $\sim 1/N$.

In Sup. Figure 3A, we simulate the classical EOMs (105) and (106) with $N_{\text{traj}} = 5000$. This large number of trajectory was chosen to ensure that the error on the sample mean is negligible small. In Sup. Figure 3B, we show one of the simulations in the ensemble used to generate Sup. Figure 3A (i.e., showing a specific value of k). In order to show more fine-grained information, we plotted the in-phase quadrature $\text{Re}(\alpha_i^{(k)}(t))$ instead of $s_i^{(k)}(t)$ (which is just the sign of the former). To generate Sup. Figure 3C, the classical EOMs are simulated for problem sizes N up to $N = 50$. Here, we use for each problem instance $N_{\text{traj}} = 100$ trajectories; this was done as the larger problem sizes required intensive computation time (on the order of one day of computation per trajectory for $N = 50$). As in the quantum simulations, the success probability used in the histogram is defined as the success probability at the end of the computation $P_{\text{success}} = P_{\text{success}}(T)$.

Supplementary Note 7. Analysis of decoherence and heating in the bus resonator

In our dynamical-coupling scheme for quantum annealing, a central bus resonator is used to establish all-to-all coupling among detuned oscillators (e.g., JPOs) while phase modulation is used to cause those oscillators to couple dynamically via the bus. It is natural to ask whether the bus can act as a non-trivial decoherence channel for the annealer. Intuitively, if there is a significant flow of excitations (i.e., photons) from the oscillators into the bus as a result of the modulations, then photon losses in the bus resonator might cause decoherence in the quantum annealer, which in turn implies that the quality of the bus resonator could be relevant to the scalability of our scheme. To better understand this “bus heating” effect and the impact of decoherence in light of it, we revisit in this section our use of the Schrieffer-Wolff expansion in deriving the native all-to-all couplings and consider explicitly the photon population in the bus and the resulting requirements on the bus’s coherence time. We find that the coherence-time requirement of the bus scales linearly with problem size, but that the requirement is not a major obstacle for the experimental realization of the dynamical-coupling scheme because the coherence-time values required for sizes up to $N = 1000$ should be easily achieved. The intuitive reason for why the bus decoherence is not a significant factor in the design of a quantum annealer using our scheme is that the bus does not get populated significantly due to its strong detuning from the oscillators, and the potential heating due to the phase modulation is on the same order as the usual heating of a bus coupled to oscillators detuned from one another.

In performing the Schrieffer-Wolff expansion that led to the coupling Hamiltonian (44), we applied a frame transformation to “polariton” modes $\hat{a}_{\text{pol } i}$ (which we approximate as $\hat{a}_i$, the lab-frame oscillator modes, in (44)) and $\hat{b}_{\text{pol}}$. The latter polariton-bus mode can be written in terms of the lab-frame modes as

$$\hat{b}_{\text{pol}} = \hat{b} + \sum_i \frac{g_i}{\Delta_i(t)} \exp\left(i \int_0^t \Delta_i(t') dt'\right) \hat{a}_i. \quad (111)$$

The time-dependent bus-oscillator detuning $\Delta_i(t) = \omega_{\text{bus}} - \omega_i(t)$ can be well approximated by $\Delta_{\text{nom } i}t - \delta\phi_i(t)$ as the fast parametric-driving terms $\omega_{\text{fast } i}(t)$ can be neglected while (54) can be used to approximate $\omega_{\text{dc } i} \approx \omega_{\text{bus}} - \Delta_{\text{nom } i}$ and $\omega_{\text{slow } i}(t) \approx \delta\phi_i(t)$. Here, $\Delta_{\text{nom } i}$ as defined by (64) is the nominal average detuning of oscillator i and $\delta\phi_i(t) = -\sum_{k=1}^{N-1} F_i^{(k)} \cos(k\Lambda t)k\Lambda$ is the instantaneous frequency shift due to the modulations for dynamical coupling. Together, these approximations allow us to write $\hat{b}$ as

$$\hat{b} = \hat{b}_{\text{pol}} - \sum_i \frac{g_i}{\Delta_i(t)} \exp[i(\Delta_{\text{nom } i}t - \delta\phi_i(t))] \hat{a}_i. \quad (112)$$

After the Schrieffer-Wolff expansion, the Hamiltonian (44) does not induce excitations of the polariton-bus mode, so $\langle \hat{b}_{\text{pol}}^\dagger \hat{b}_{\text{pol}} \rangle = 0$. Thus, the energy in the lab-frame bus mode is strictly due to its hybridization with the lab-frame oscillator modes, and this population is given by

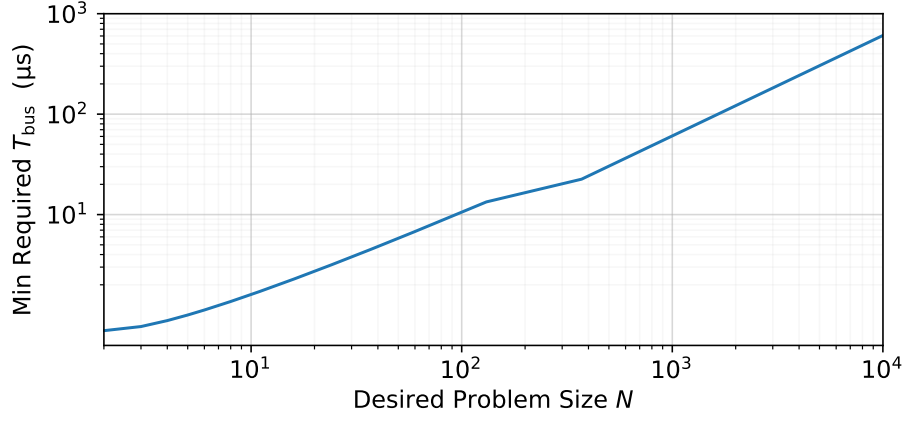
$$\langle \hat{b}^\dagger \hat{b} \rangle = \sum_i \frac{g_i^2}{\Delta_i^2(t)} \langle \hat{a}_i^\dagger \hat{a}_i \rangle + \sum_{i \neq j} \frac{g_i g_j}{\Delta_i(t) \Delta_j(t)} \exp\left[-i(\Lambda(i-j)t + \delta\phi_j(t) - \delta\phi_i(t))\right] \langle \hat{a}_i^\dagger \hat{a}_j \rangle, \quad (113)$$

since $\Delta_{\text{nom } i} - \Delta_{\text{nom } j} = (i-j)\Lambda$. The first term is the usual heating of a bus coupled to oscillators detuned from one another, while the second term is unique to our dynamical coupling scheme and is due to interference of the sidebands that each oscillator has resulting from the modulations of the oscillators. The second term can be simplified by noting that time average of the phase term is related to the effective coupling matrix C_{eff} . If we further assume that over a single Floquet period, we can approximate the factor $g_i/\Delta_i(t)$ as a constant g_i/Δ_i in the time-averaging integral, then the bus photon number averaged over the Floquet period is given by

$$\langle \hat{b}^\dagger \hat{b} \rangle_{\text{avg}} = \sum_i \frac{g_i^2}{\Delta_i^2} \langle \hat{a}_i^\dagger \hat{a}_i \rangle + \sum_{i \neq j} \frac{g_i g_j}{\Delta_i \Delta_j} \frac{\lambda_C}{J_{ij}} C_{\text{eff}} \langle \hat{a}_i^\dagger \hat{a}_j \rangle. \quad (114)$$

To arrive at a further simplification, suppose we treat the factor g_i/Δ_i as well as the oscillator photon numbers $\langle \hat{a}_i^\dagger \hat{a}_i \rangle$ to be approximately uniform over all oscillators, so that $g_i/\Delta_i \approx g/\Delta$ and $\langle \hat{a}_i^\dagger \hat{a}_i \rangle \approx n$. A consequence of the latter approximation is that $\frac{1}{2} \langle \hat{a}_i^\dagger \hat{a}_j + \hat{a}_j^\dagger \hat{a}_i \rangle \approx n\sigma_i\sigma_j$, where $\sigma_i \in \{+1, -1\}$ are the spin configurations encoded by the oscillators. Additionally, suppose that the fidelity of the effective coupling matrix is high, so that $C_{\text{eff } ij} \approx C_{ij}$. Then the population in the bus takes the form

$$\langle \hat{b}^\dagger \hat{b} \rangle_{\text{avg}} \approx Nn \frac{g^2}{\Delta^2} \left(1 + \frac{\eta}{N} \sum_{i \neq j} C_{ij} \sigma_i \sigma_j \right), \quad (115)$$



Supplementary Figure 14. Scaling requirements of bus resonator coherence times. Minimum required coherence time T_{bus} of the bus as given by (117) as a function of N (for SK7 problems). The various parameters ($\tilde{r}_{\text{max}}, \tilde{\lambda}_J, \Delta_{\text{nom1}}$) are chosen in accordance with the prescription specified in Sup. Note 5 A.

where we note that the sideband contribution depends essentially on the Ising energy $\sum_{i \neq j} C_{ij} \sigma_i \sigma_j$. A crucial point in interpreting this equation is that, while the scaling of the summation goes as $\mathcal{O}(N^2)$, η scales as $\mathcal{O}(1/N)$, so the overall scaling of the term in the brackets is $\mathcal{O}(1)$; this means that the sideband interferences, even in the worst case, can only cause heating comparable with the level expected in a bus coupled to oscillators detuned from one another. Furthermore, the oscillator system has been constructed to operate as a quantum annealer that minimizes the Ising energy of a given Ising problem, so the sum $\sum_{i \neq j} C_{ij} \sigma_i \sigma_j$ will be driven to have a low value [1]. In fact, if the Ising energy of the Ising configuration encoded by the oscillators is negative, the sidebands will actually *cool* the bus.

We now turn to quantifying the limitations that bus decoherence impose on the design of a dynamically-coupled quantum annealer. We model the decoherence of the bus with a Lindblad operator $\sqrt{\kappa_{\text{bus}}} \hat{b}$, where κ_{bus} represents the photon loss rate of the bus. To ensure that the annealer is unaffected by the decoherence of the bus, we require that $\kappa_{\text{bus}} \langle \hat{b}^\dagger \hat{b} \rangle \ll \chi$. Using $g_i^2 / \Delta_i = \lambda_J$, $n \approx r / \chi \lesssim \tilde{r}_{\text{max}}$, and $\Delta_i \approx \Delta_{\text{nom1}} \gtrsim \Delta_{\text{nom1}}$, this requirement can be stated as

$$\kappa_{\text{bus}} \ll \frac{\Delta_{\text{nom1}}}{N \tilde{r}_{\text{max}} \tilde{\lambda}_J}. \quad (116)$$

This requirement on the bus decay rate can be recast as a requirement on the bus-photon lifetime $T_{\text{bus}} = 1 / \kappa_{\text{bus}}$. Considering $\kappa_{\text{bus}} \langle \hat{b}^\dagger \hat{b} \rangle \leq \chi / 5$ for concreteness, the photon lifetime of the bus should obey

$$T_{\text{bus}} \geq 5 \frac{N \tilde{r}_{\text{max}} \tilde{\lambda}_J}{\Delta_{\text{nom1}}}, \quad (117)$$

which we take to be the minimum required bus-photon lifetime to avoid decoherence due to bus photon losses. The nominal detuning Δ_{nom1} of the first oscillator is constant at large N , since our scaling scheme is detuning-limited at large N . In addition, $\tilde{r}_{\text{max}}$ and $\tilde{\lambda}_J$ are also approximately constant ($\tilde{\lambda}_J$ is constant for SK7 problems and logarithmic for MAX-CUT). Thus, we conclude that the minimum required coherence time of the bus at large N scales as N , in the same way as the coherence time T_{cav} of the oscillator qubits (and essentially for the same reason, namely that χ scales as $1/N$). In Sup. Figure 16, we plot the minimum required T_{bus} as a function of N , and observe the expected linear scaling at large N . Furthermore, for moderate problem sizes ($N < 100$), we see that coherence times $T_{\text{bus}} \geq 10 \mu\text{s}$ are sufficient. Coplanar waveguide resonators (of the kind used in bus architectures) have experimentally-measured coherence times of $100 \mu\text{s}$ [18, 19], which indicates that the coherence-time requirements for moderate problem sizes can be easily met and also suggests the requirements for large N may not be prohibitive.

To numerically study the effects of bus loss and verify some of the arguments made above, we perform quantum simulations on a full Hamiltonian that includes both the oscillator and bus modes (i.e., not using the Schrieffer-Wolff expansion to eliminate the bus mode). The full Hamiltonian in the lab frame is given by combining (39) with (41); if we write this Hamiltonian in a rotating frame where the oscillators have a detuning $-(\delta_i - g_i^2 / \Delta_i(t))$, the full Hamiltonian becomes

$$\sum_i \left[- \left(\delta_i - \frac{g_i^2}{\Delta_i(t)} \right) \hat{a}_i^\dagger \hat{a}_i - \frac{\chi}{2} \hat{a}_i^{\dagger 2} \hat{a}_i^2 + \left(\frac{r(t)}{2} \hat{a}_i^2 + \text{H.c.} \right) + \left(g_i \exp \left[-i \int_0^t \left(\Delta_i(t') - \delta_i + \frac{g_i^2}{\Delta_i(t')} \right) dt' \right] \hat{a}_i^\dagger \hat{b} + \text{H.c.} \right) \right],$$

where we assume that the $\omega_{\text{fast}i}(t)$ component of $\omega_{\text{ac}i}(t)$ is chosen appropriately to produce the pump coefficient $r(t)$.

To simplify our simulations, we first make the additional assumption that $g_i^2/\Delta_i(t) \approx g_i^2/\Delta_{\text{nom}i}$, which makes the first term in the Hamiltonian time-independent. Then to simplify the phase factor in the last term, we turn to (52), which tells us that the flux modulations have been designed so that

$$\omega_i(t) - \frac{g_i^2}{\Delta_i(t)} + \delta_i \approx \omega_{\text{bus}} - \Delta_{\text{nom}i} + \delta\dot{\phi}_i(t), \quad (118)$$

having made the assumption on the left-hand side that $\omega_i(t) \approx \omega_{\text{dc}i} + \omega_{\text{slow}i}(t)$ (again, by simply neglecting the fast parametric-driving part); the right-hand side follows from (64) and the definition of $\delta\dot{\phi}_i(t)$. This equation then directly implies that the integrand inside the phase factor is well-approximated by $\Delta_{\text{nom}i} - \delta\dot{\phi}_i(t)$. Putting this all together, the Hamiltonian we simulate is

$$\sum_i \left[- \left(\delta_i - \frac{g_i^2}{\Delta_{\text{nom}i}} \right) \hat{a}_i^\dagger \hat{a}_i - \frac{\chi}{2} \hat{a}_i^{\dagger 2} \hat{a}_i^2 + \left(\frac{r(t)}{2} \hat{a}_i^2 + \text{H.c.} \right) + \left(g_i \exp[-i(\Delta_{\text{nom}i}t - \delta\phi_i(t))] \hat{a}_i^\dagger \hat{b} + \text{H.c.} \right) \right]. \quad (119)$$

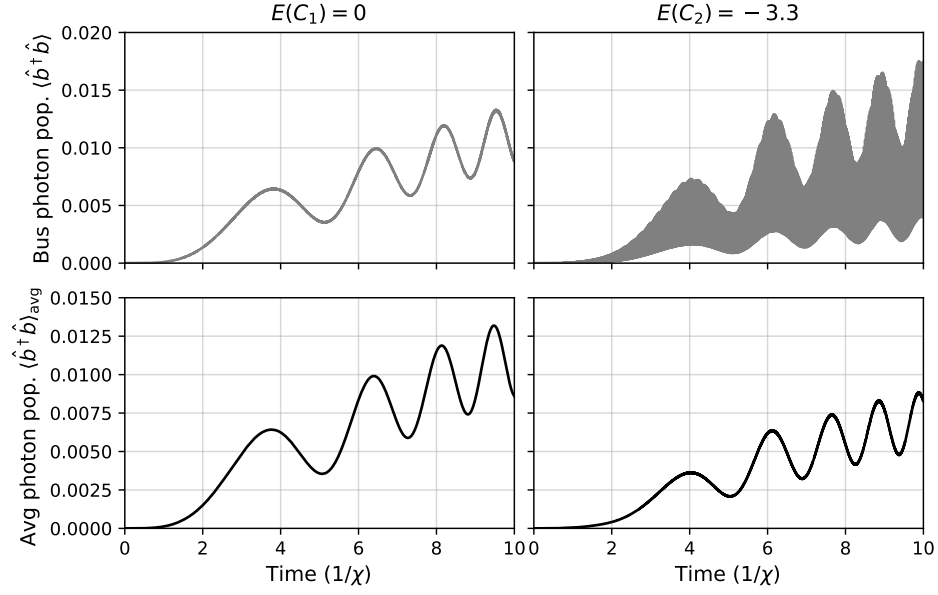
This Hamiltonian approximates the full Hamiltonian (and hence its Schrieffer-Wolff expansion approximates the dynamical-coupling Hamiltonian (2) in the main text) when the bus-oscillator detunings $\Delta_i(t) \approx \Delta_{\text{nom}i}$. To ensure that the quantum simulations we run in this section lie within this limit, we study specific Ising problems whose corresponding phase-modulation coefficients $F_k^{(i)}$ are small. (This latter stipulation can be removed if we directly insert the time-dependence of $\Delta_i(t)$ by solving (52) rather than approximating it with $\Delta_{\text{nom}i}$ in the first term of the Hamiltonian, but we make this approximation here for the sake of simplicity.)

In Sup. Figure 15, we use (119) to examine the population of the bus resonator for the following two Ising problems, for the lossless case of $\kappa_{\text{bus}} = \kappa = 0$:

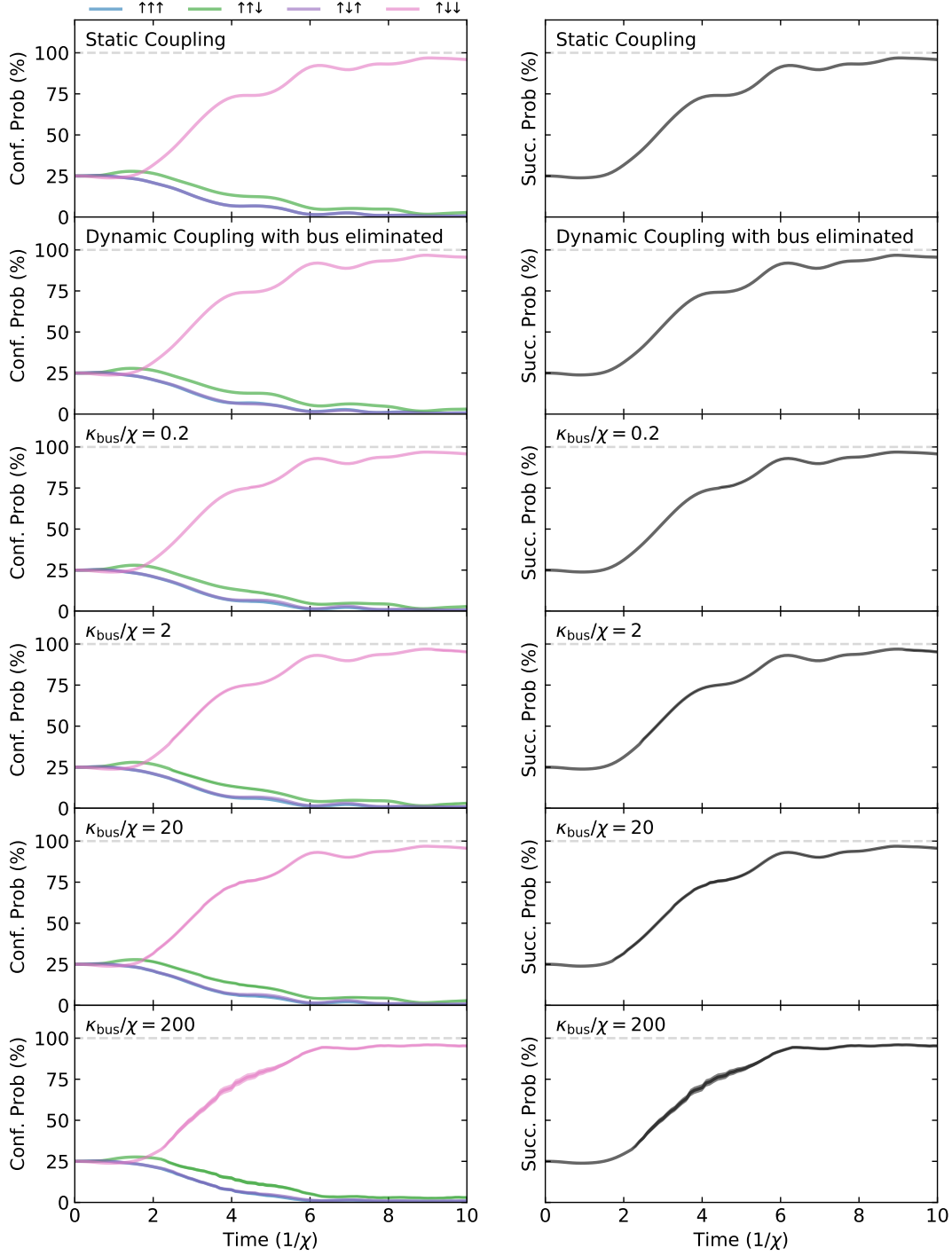
$$C_1 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad C_2 := \begin{bmatrix} 0 & 1/3 & 1 \\ 1/3 & 0 & -1/3 \\ 1 & -1/3 & 0 \end{bmatrix}. \quad (120)$$

The first Ising problem has no spin-spin couplings; when run on a dynamically-coupled quantum annealer there will be no phase modulations applied to the oscillators, and consequently the oscillators will be uncoupled because they are detuned from one another. Consistent with (113), we observe that the bus population (113) is generally small with a maximum value of about 1% of a single photon, and there are no fast oscillations of the population in time as no phase modulations are applied. On the other hand, when the dynamically-coupled quantum annealer is run with problem C_2 , the bus population exhibits fast oscillations in time due to the sidebands from the oscillator modulations. Consistent with (115), we observe that the bus's average photon population is smaller than in the case of C_1 as the ground state energy $E(C_2)$ of the problem C_2 is negative.

In Sup. Figure 16, we see the results of simulating (119) in the presence of bus decoherence, for the case where the oscillators are lossless. Specifically, we investigate the effect of the bus decoherence on the performance of the quantum annealer by comparing its success and configuration probabilities to the corresponding statically-coupled and dynamically-coupled system with the bus mode eliminated. We find that even for high bus photon-decay rates of $\kappa_{\text{bus}} = 200\chi$, both the success and configuration probabilities are qualitatively similar to that of the statically-coupled system, showing that even for high bus photon-decay rates the dynamically-coupled quantum annealer performs as desired. These results are consistent with the fact that the bus is very weakly populated as shown in Sup. Figure 15.



Supplementary Figure 15. Population of the bus for different Ising problems. Top row: The expectation value of the photon number $\langle \hat{b}^\dagger \hat{b} \rangle$ in the bus resonator as a function of time for two different Ising problem matrices C_1, C_2 with respective ground state energies $E(C_1), E(C_2)$. Bottom row: The time average of $\langle \hat{b}^\dagger \hat{b} \rangle$ over the Floquet timescale $1/\Lambda$ for each corresponding Ising problem. The quantum simulations are performed using the Hamiltonian (119). Here, decoherence in both the bus and the oscillators are assumed to be zero ($\kappa_{\text{bus}} = \kappa = 0$), while the system parameters are $T_{\text{ramp}}\chi = 10$, and $r_{\text{max}}/\chi = 3.5$, with $g_i, \Delta_{\text{nom}i}, J$, and Λ chosen according to the prescription given in Sup. Note 5 A.



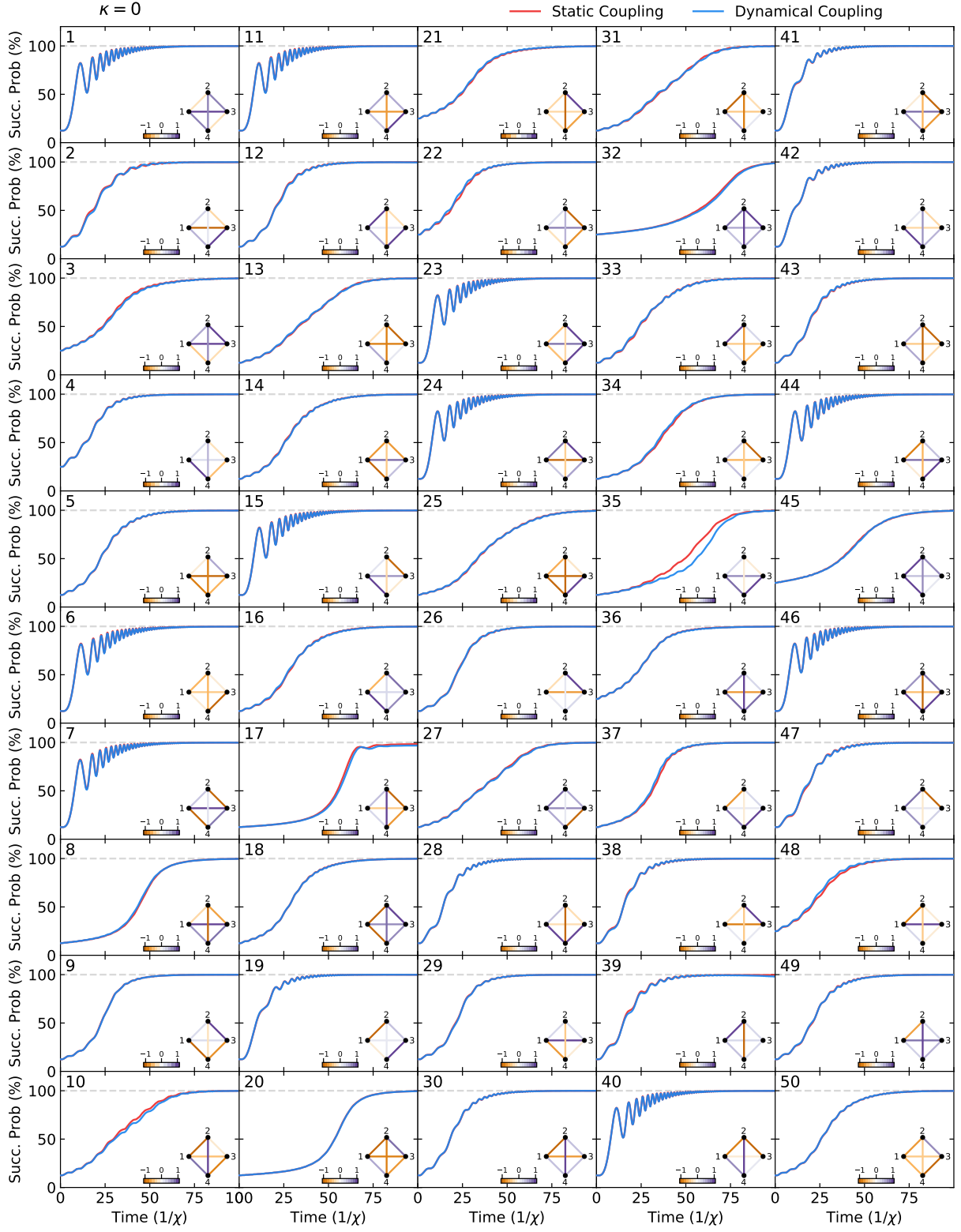
Supplementary Figure 16. Performance of dynamically-coupled quantum annealer with bus explicitly modeled, for various bus-photon-decay rates. Quantum simulations of the dynamically-coupled system model with the bus mode included as described by the Hamiltonian (119), for various values of the bus-photon-decay rate κ_{bus} . For comparison, we also simulate the statically-coupled system (1, main text) and the dynamically-coupled system with the bus mode eliminated after the Schrieffer-Wolff expansion (2, main text). Left column: The probabilities of obtaining each of the possible spin configurations (i.e., configuration probability) as a function of time. Right column: Success probabilities of finding a ground state of the Ising problem as a function of time. The Ising problem matrix used in these simulations is C_2 , and the oscillators are assumed to be lossless ($\kappa = 0$). The system parameters are $T_{\text{ramp}}\chi = 10$, and $r_{\text{max}}/\chi = 3.5$, with g_i , Δ_{nom_i} , J , and Λ chosen according to the prescription in Sup. Note 5 A.

Supplementary Note 8. Additional Data Figures

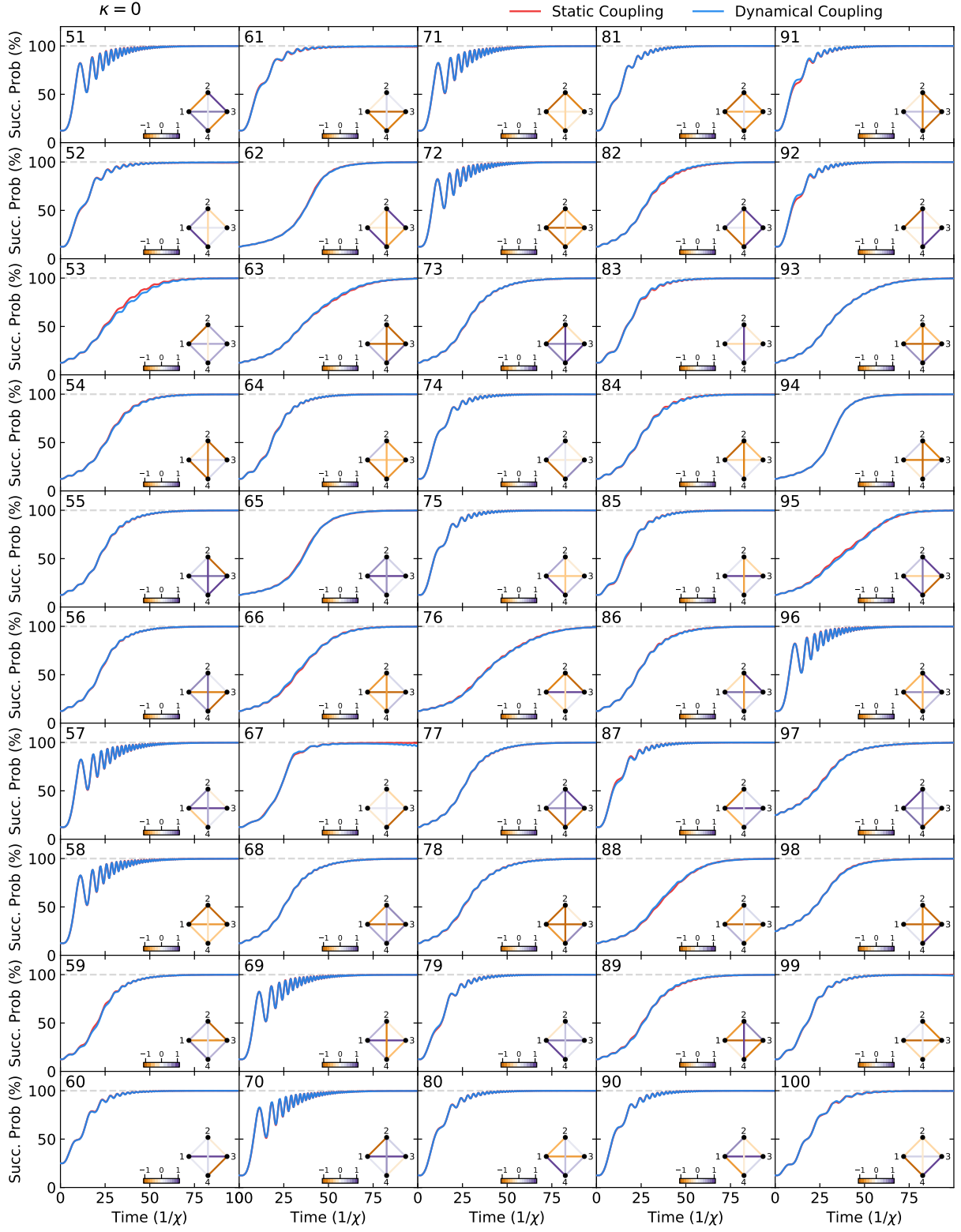
In this section, we provide additional figures which depict in greater detail the data generated from the quantum and classical simulations performed in this work. The more comprehensive statistics provided by these figures support the conclusion made in the main manuscript that the dynamical-coupling scheme results in time-evolution of success probabilities in a way which closely matches that of the corresponding statically-coupled system. They also provide a check that the figures presented in the main manuscript are statistically representative of the entire data set generated from our simulations.

In Sup. Figures 17 through 22, we compare the success probability $P_{\text{success}}(t)$ of the statically- and dynamically-coupled systems for each of the 100 problem instances we used in our quantum simulations. These figures investigate, respectively, the zero-loss, low-loss, and high-loss cases, indicating that, in most cases, the dynamical-coupling scheme produces the desired evolution, at least in distribution. Corresponding to each of the plots, Sup. Figures 23 through 28 show, for the statically-coupled system, the configuration probabilities $P_{\sigma}(t)$ for each of the spin configurations σ , while Sup. Figures 29 through 34 show the corresponding configuration probabilities for the dynamically-coupled system. Again, we see that the trajectories match well, at least in distribution.

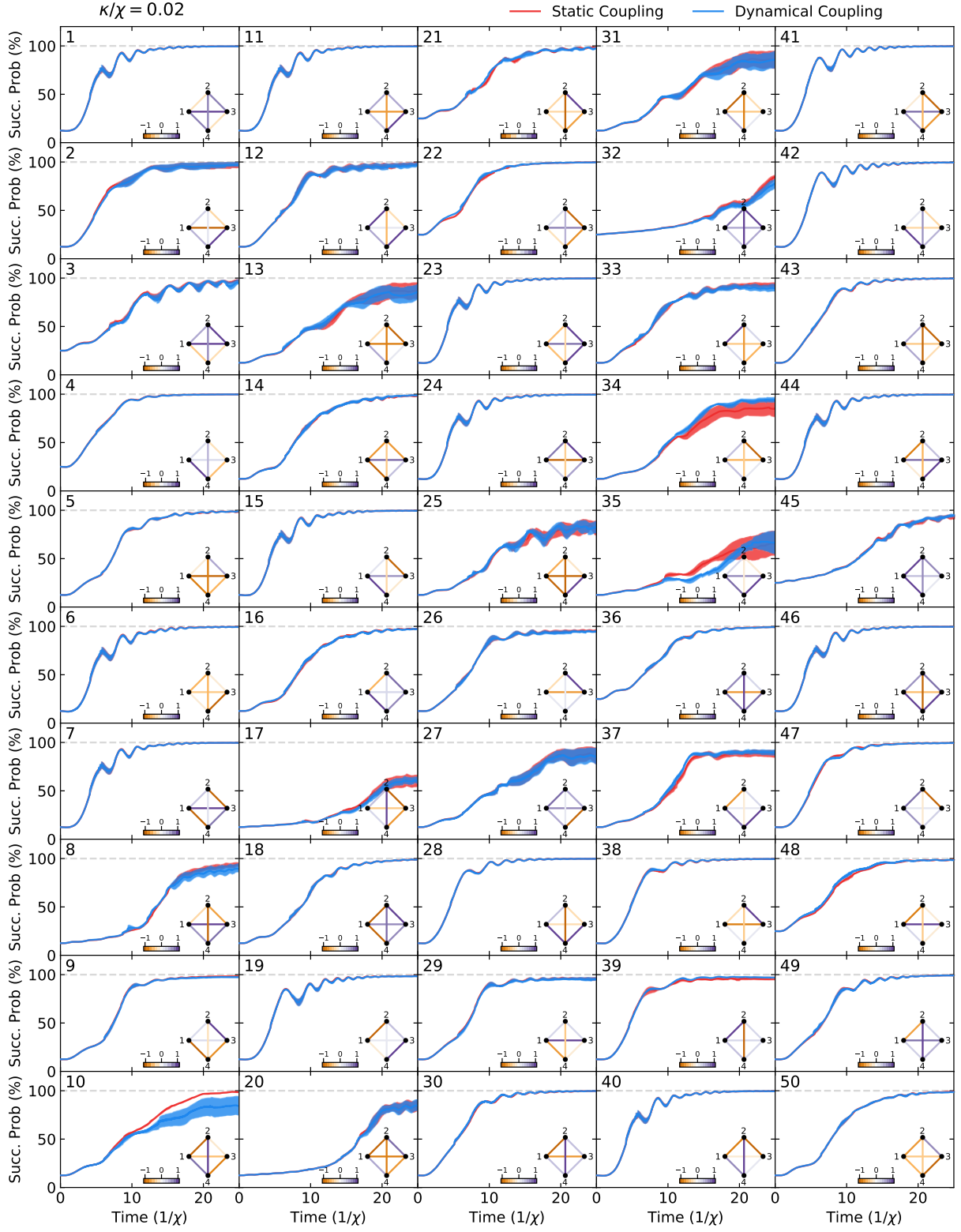
Finally, in Sup. Figure 35, we show the same correlation/histogram plots used in the bottom of Sup. Figure 3C of the main manuscript, for all the problem sizes that we have numerically simulated. Note that the projected histograms correspond to the histogram shown in Sup. Figure 3B of the main manuscript. The data shows good correspondence in the success probability between the dynamically- and statically-coupled systems for all problem sizes considered.



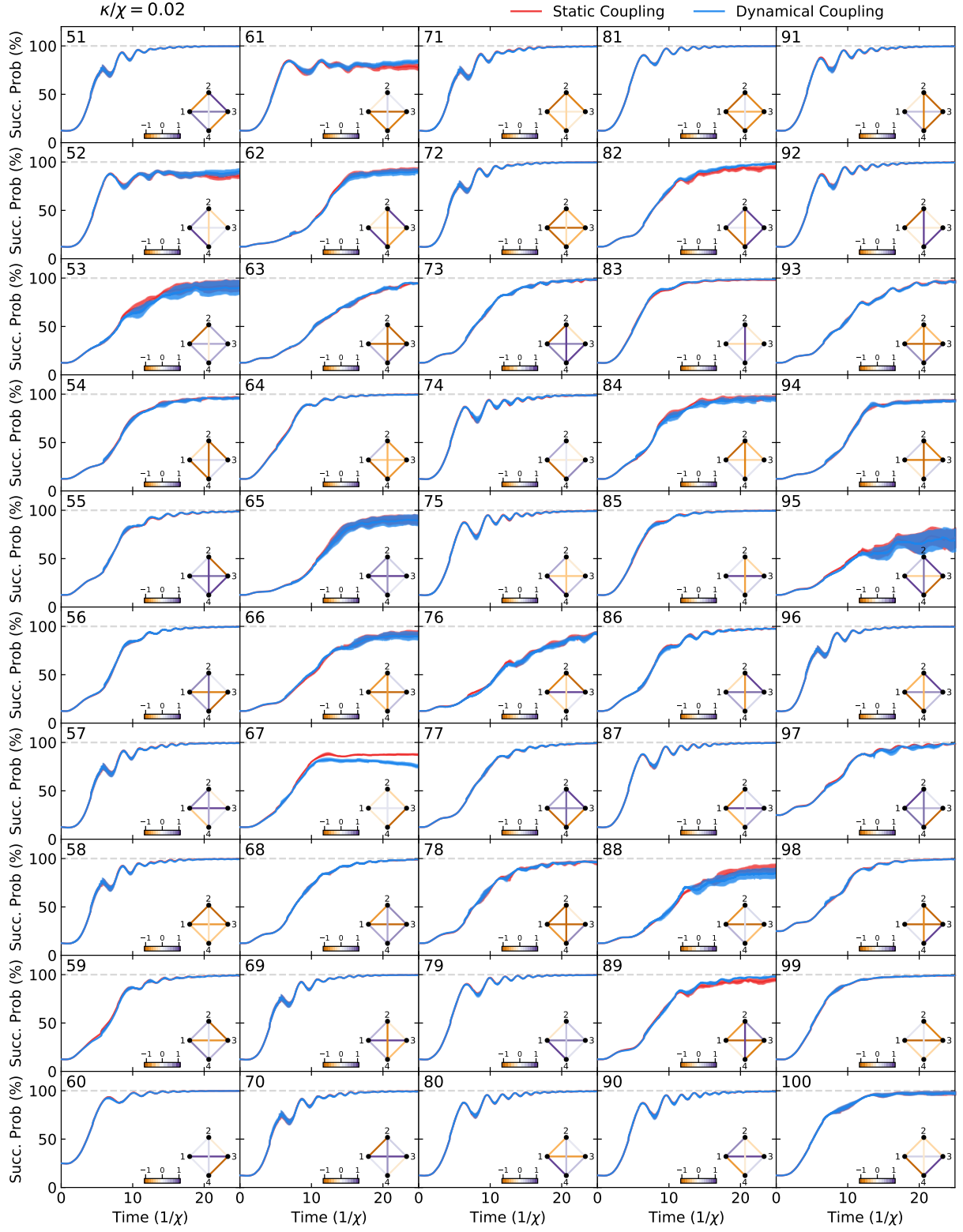
Supplementary Figure 17. Quantum simulation of success probability, lossless case (part 1/2). Evolution of the success probability $P_{\text{success}}(t)$ of finding a ground state of the inset Ising problem instance (one for each sub-axis, numbered), assuming the cavity-photon decay rate $\kappa = 0$. In red is shown the success probability of a statically-coupled system implementing the Ising couplings exactly, while in blue is shown the success probability of the corresponding dynamically-coupled system. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology.



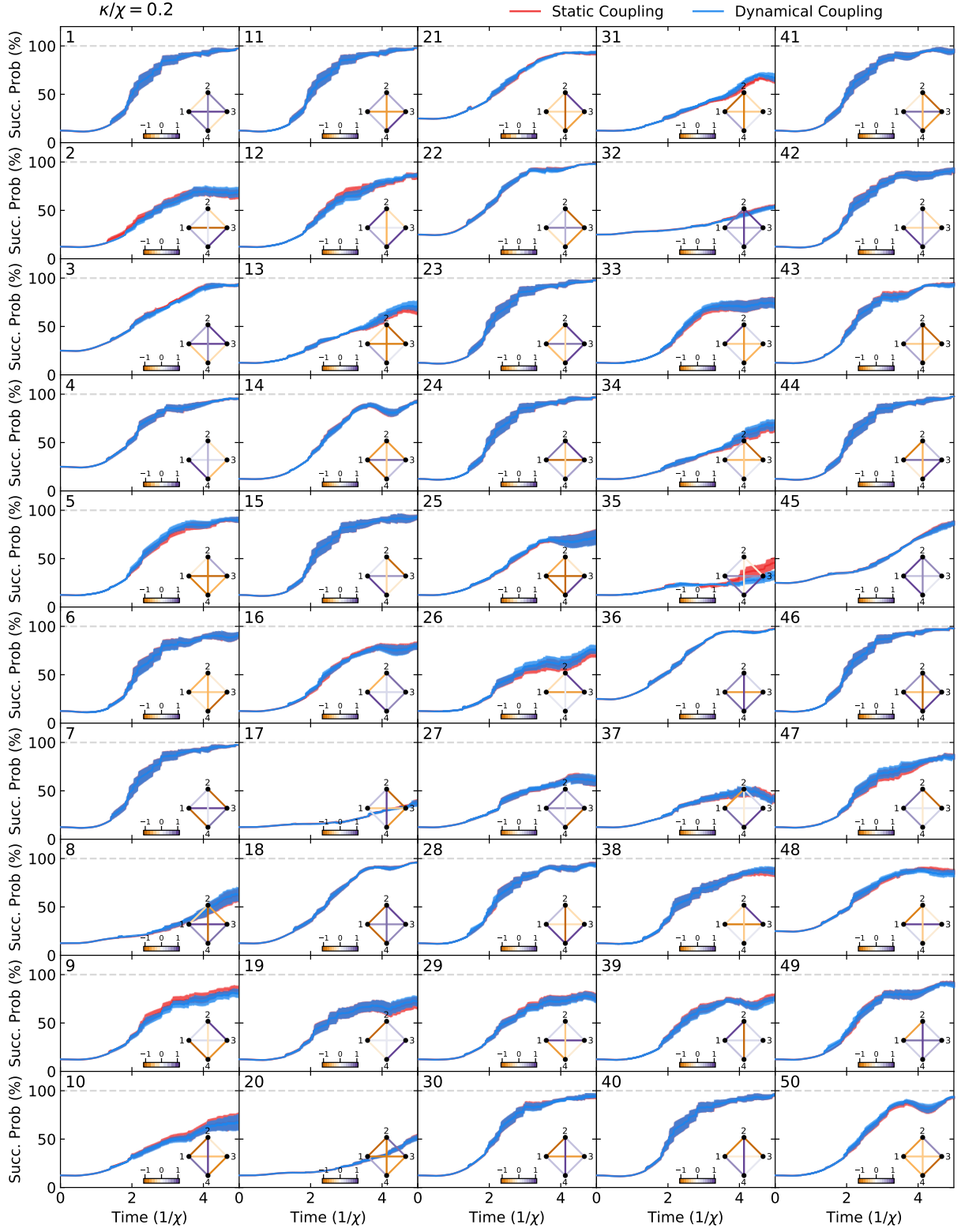
Supplementary Figure 18. Quantum simulation of success probability, lossless case (part 2/2). Evolution of the success probability $P_{\text{success}}(t)$ of finding a ground state of the inset Ising problem instance (one for each sub-axis, numbered), assuming the cavity-photon decay rate $\kappa = 0$. In red is shown the success probability of a statically-coupled system implementing the Ising couplings exactly, while in blue is shown the success probability of the corresponding dynamically-coupled system. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology.



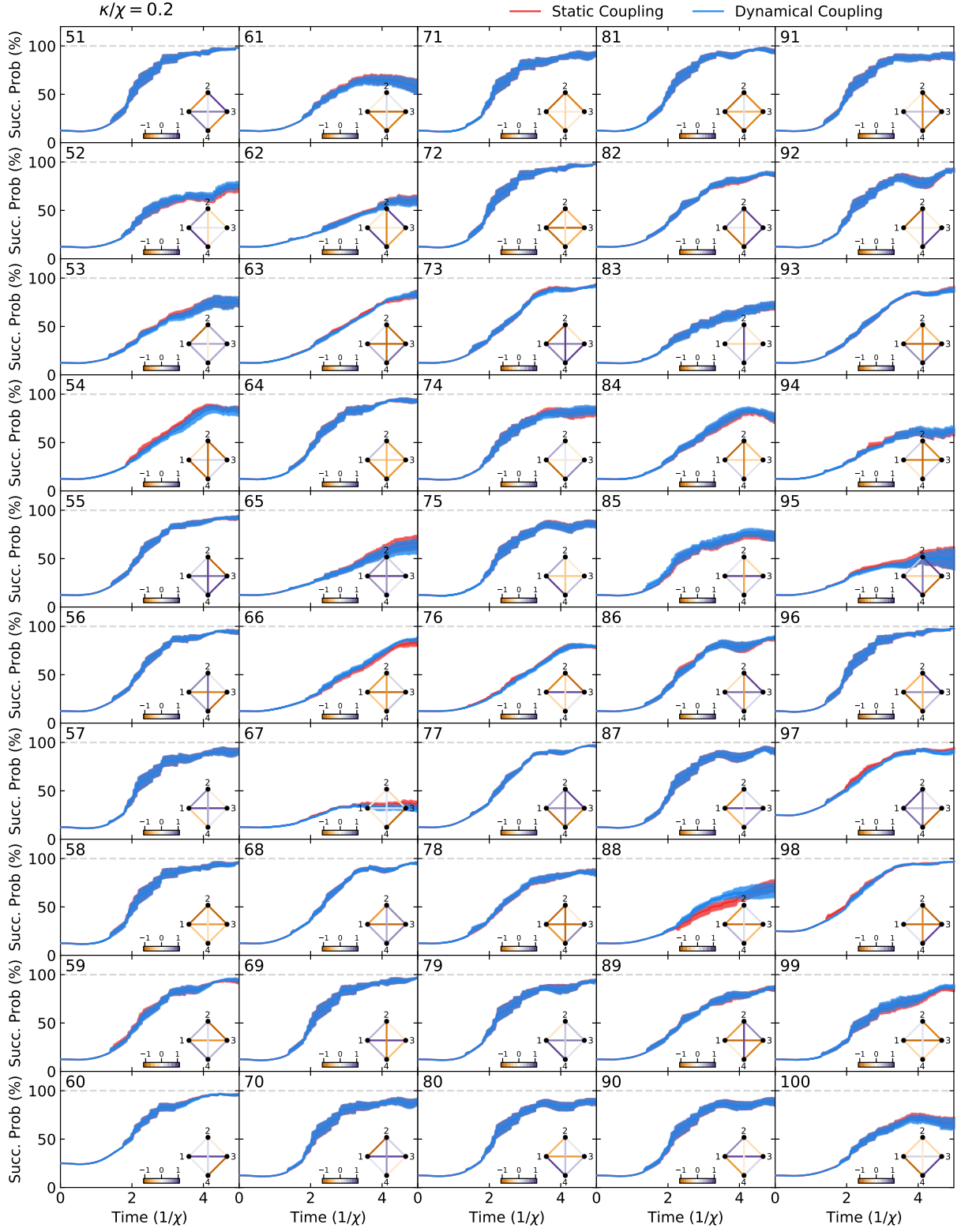
Supplementary Figure 19. Quantum simulation of success probability, low-loss case (part 1/2). Evolution of the success probability $P_{\text{success}}(t)$ of finding a ground state of the inset Ising problem instance (one for each sub-axis), assuming the cavity-photon decay rate $\kappa = 0.02\chi$. In red is shown the success probability of a statically-coupled system implementing the Ising couplings exactly, while in blue is shown the success probability of the corresponding dynamically-coupled system. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



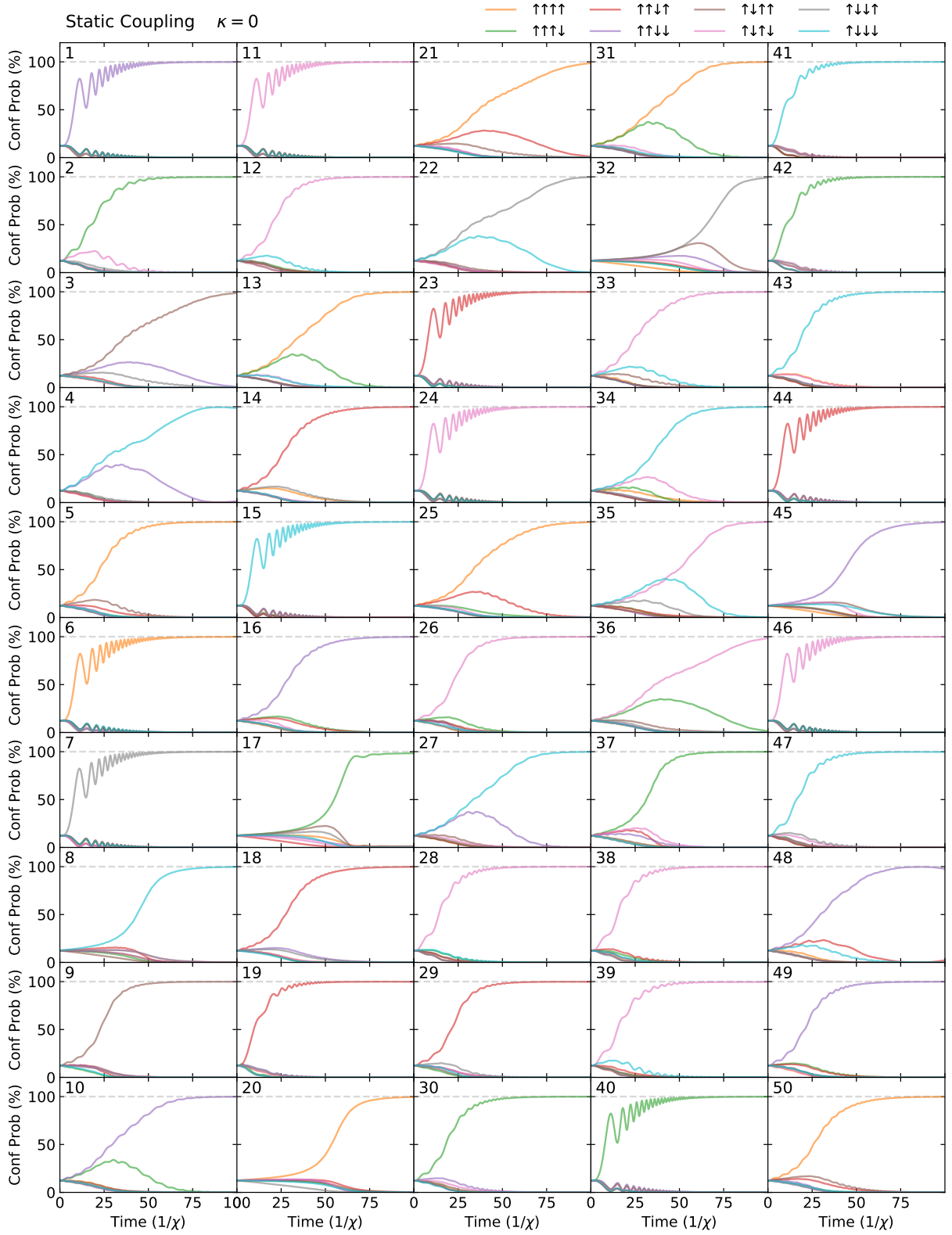
Supplementary Figure 20. Quantum simulation of success probability, low-loss case (part 2/2). Evolution of the success probability $P_{\text{success}}(t)$ of finding a ground state of the inset Ising problem instance (one for each sub-axis), assuming the cavity-photon decay rate $\kappa = 0.02\chi$. In red is shown the success probability of a statically-coupled system implementing the Ising couplings exactly, while in blue is shown the success probability of the corresponding dynamically-coupled system. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



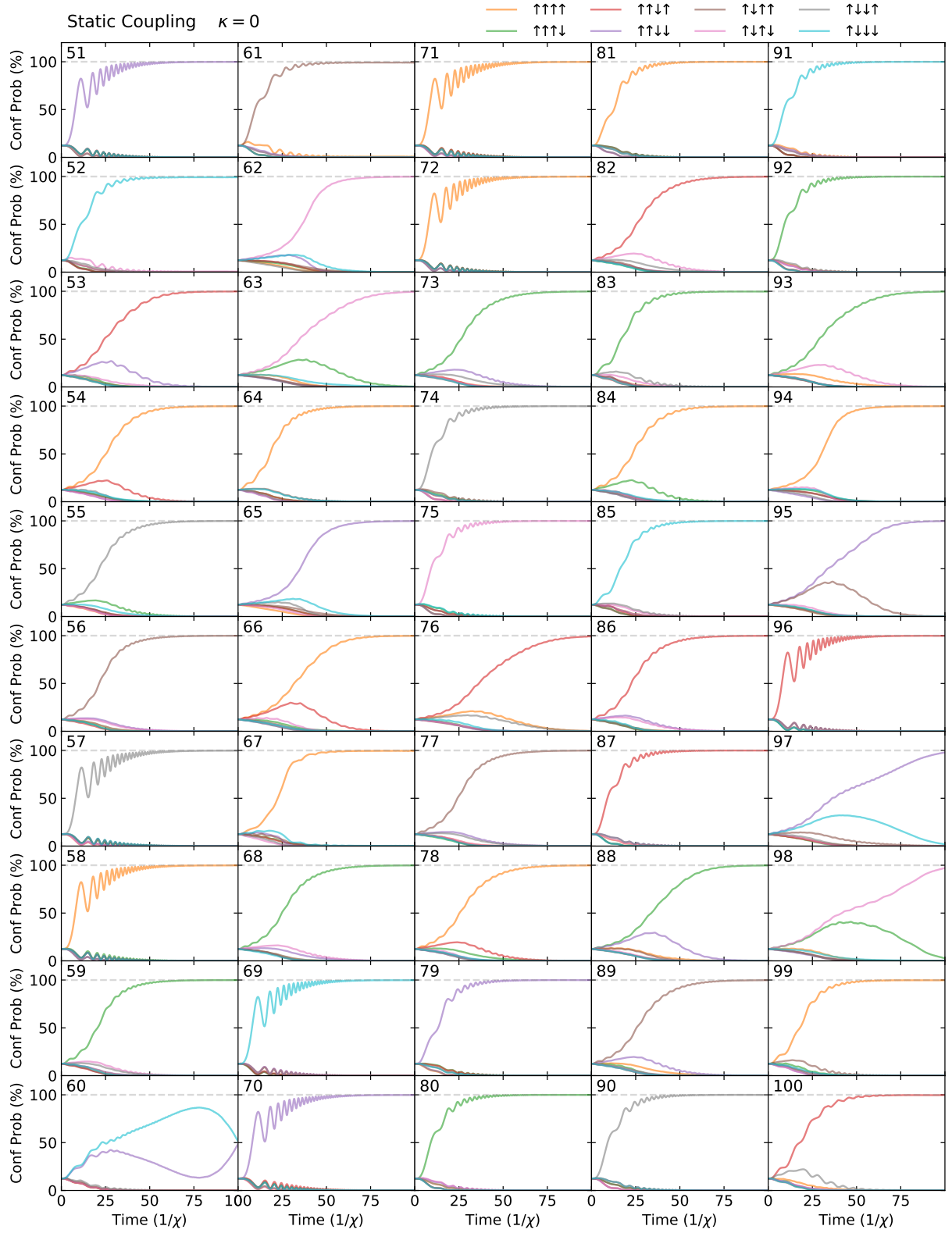
Supplementary Figure 21. Quantum simulation of success probability, high-loss case (part 1/2). Evolution of the success probability $P_{\text{success}}(t)$ of finding a ground state of the inset Ising problem instance (one for each sub-axis), assuming the cavity-photon decay rate $\kappa = 0.2\chi$. In red is shown the success probability of a statically-coupled system implementing the Ising couplings exactly, while in blue is shown the success probability of the corresponding dynamically-coupled system. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



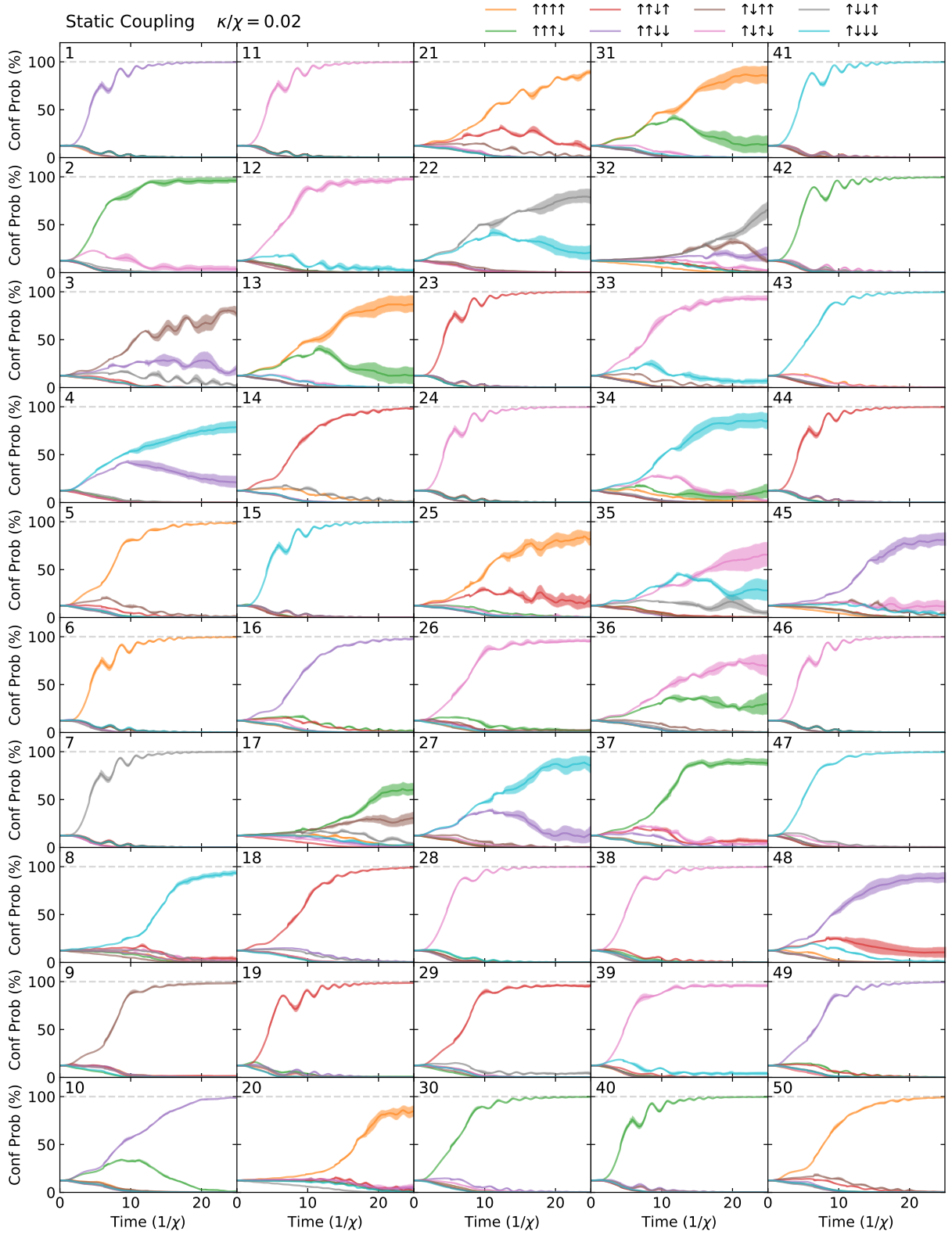
Supplementary Figure 22. Quantum simulation of success probability, high-loss case (part 2/2). Evolution of the success probability $P_{\text{success}}(t)$ of finding a ground state of the inset Ising problem instance (one for each sub-axis), assuming the cavity-photon decay rate $\kappa = 0.2\chi$. In red is shown the success probability of a statically-coupled system implementing the Ising couplings exactly, while in blue is shown the success probability of the corresponding dynamically-coupled system. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



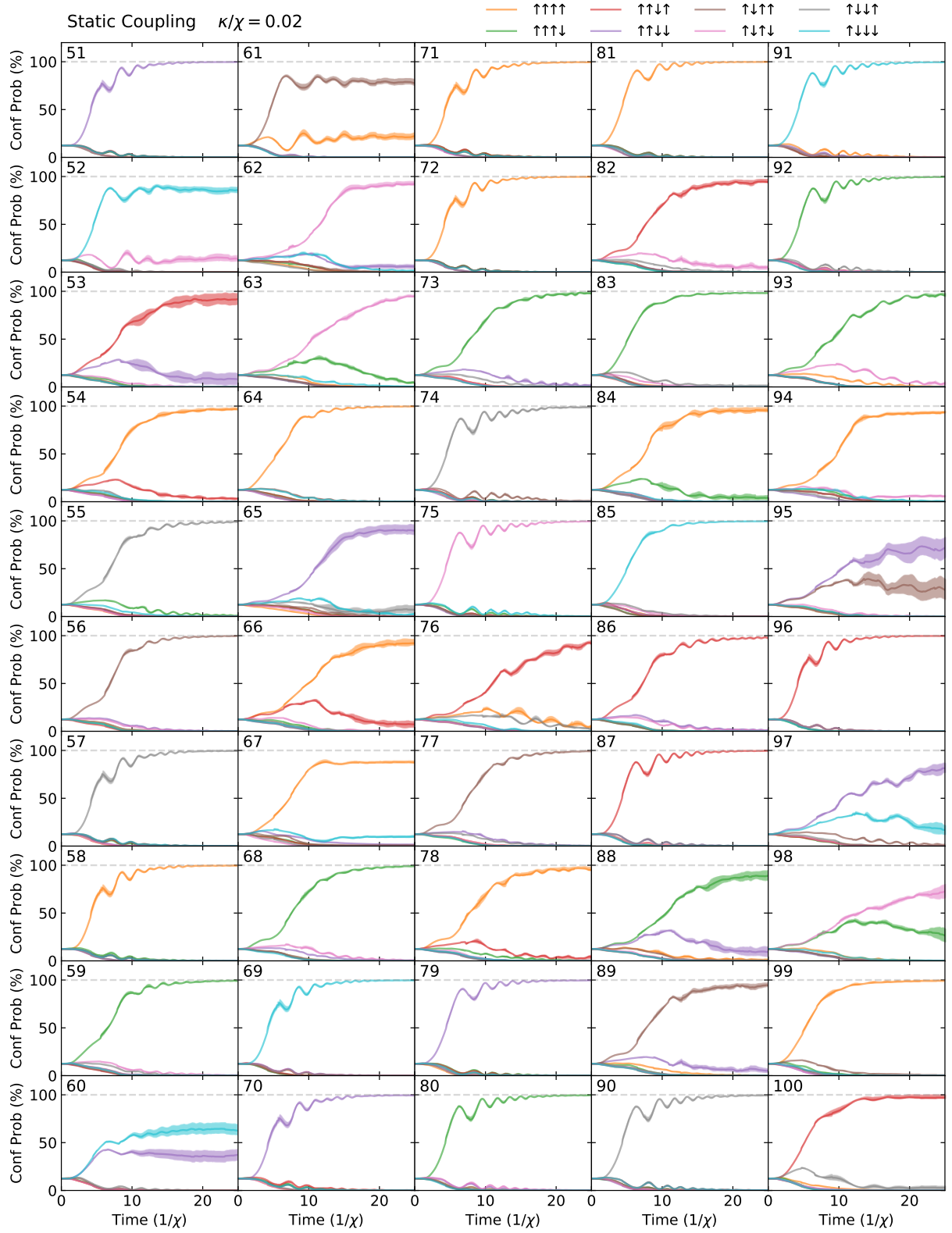
Supplementary Figure 23. Configuration probabilities of statically-coupled system, lossless case (part 1/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the statically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 17. The configuration probability is defined according to 93 in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology.



Supplementary Figure 24. Configuration probabilities of statically-coupled system, lossless case (part 2/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the statically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 18. The configuration probability is defined according to 93 in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology.

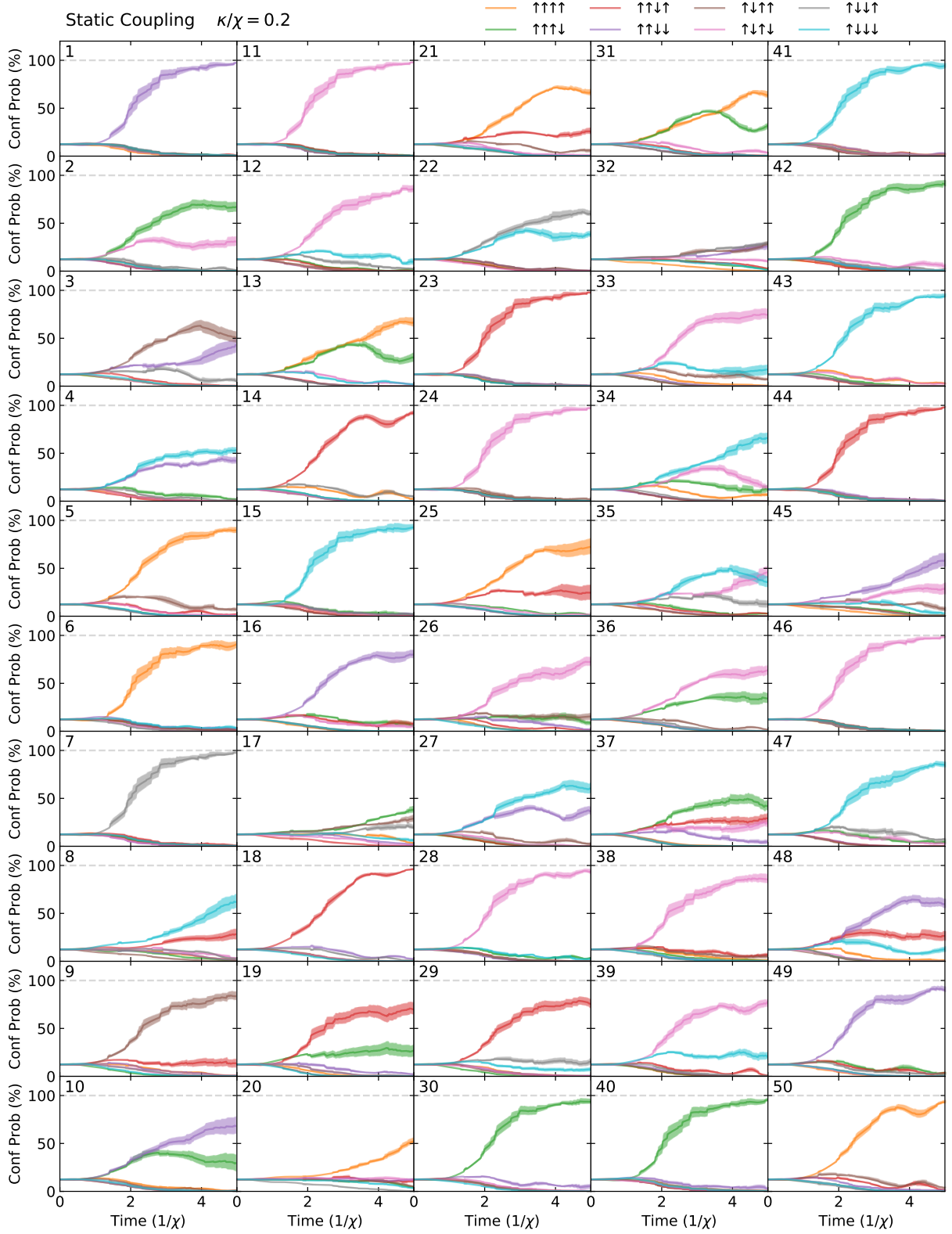


Supplementary Figure 25. Configuration probabilities of statically-coupled system, low-loss case (part 1/2). Evolution of the probabilities $P_\sigma(t)$ of obtaining each of the possible spin configurations in the statically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 17. The configuration probability is defined according to 93 in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.

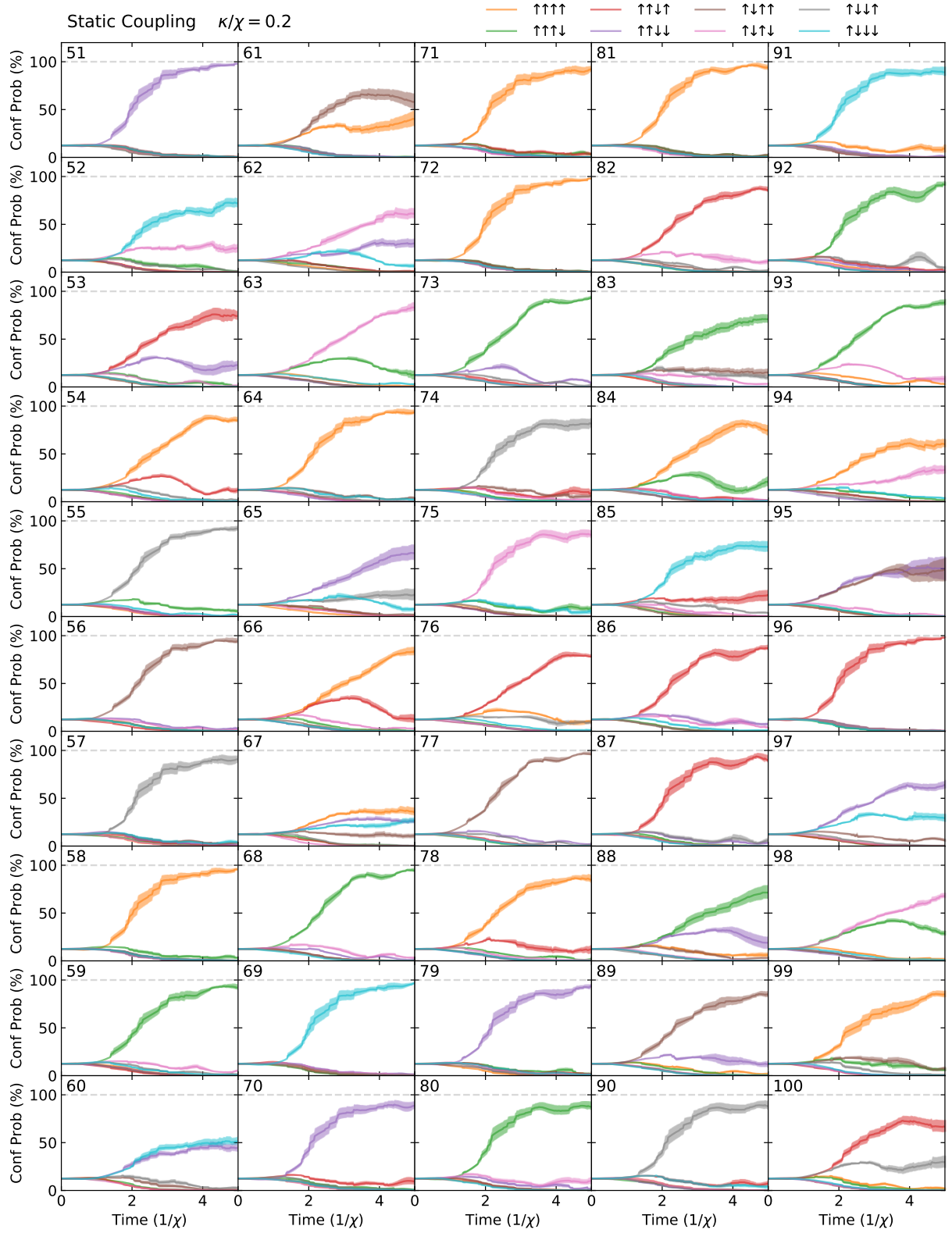


Supplementary Figure 26. Configuration probabilities of statically-coupled system, low-loss case (part 2/2).

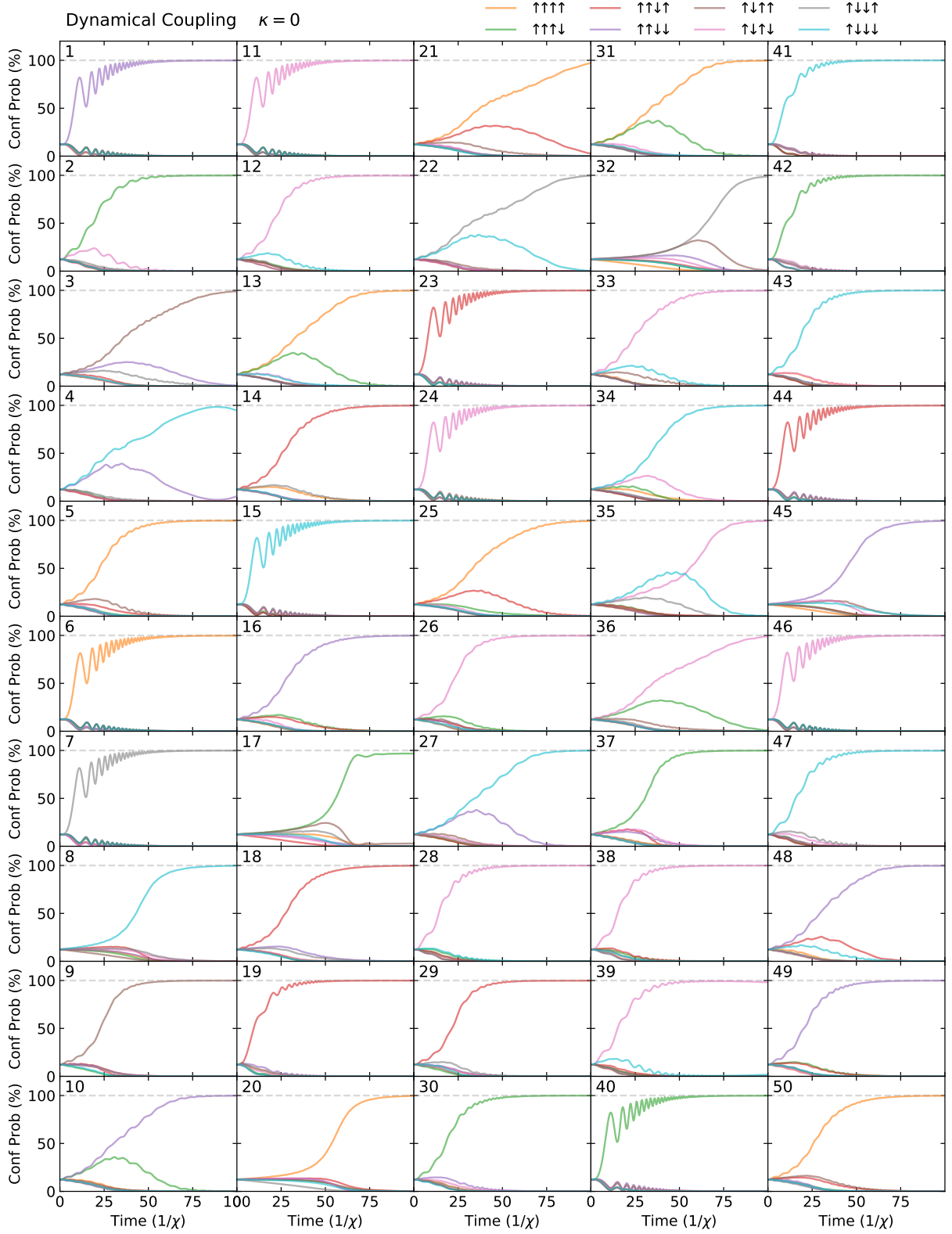
Evolution of the probabilities $P_\sigma(t)$ of obtaining each of the possible spin configurations in the statically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 18. The configuration probability is defined according to 93 in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



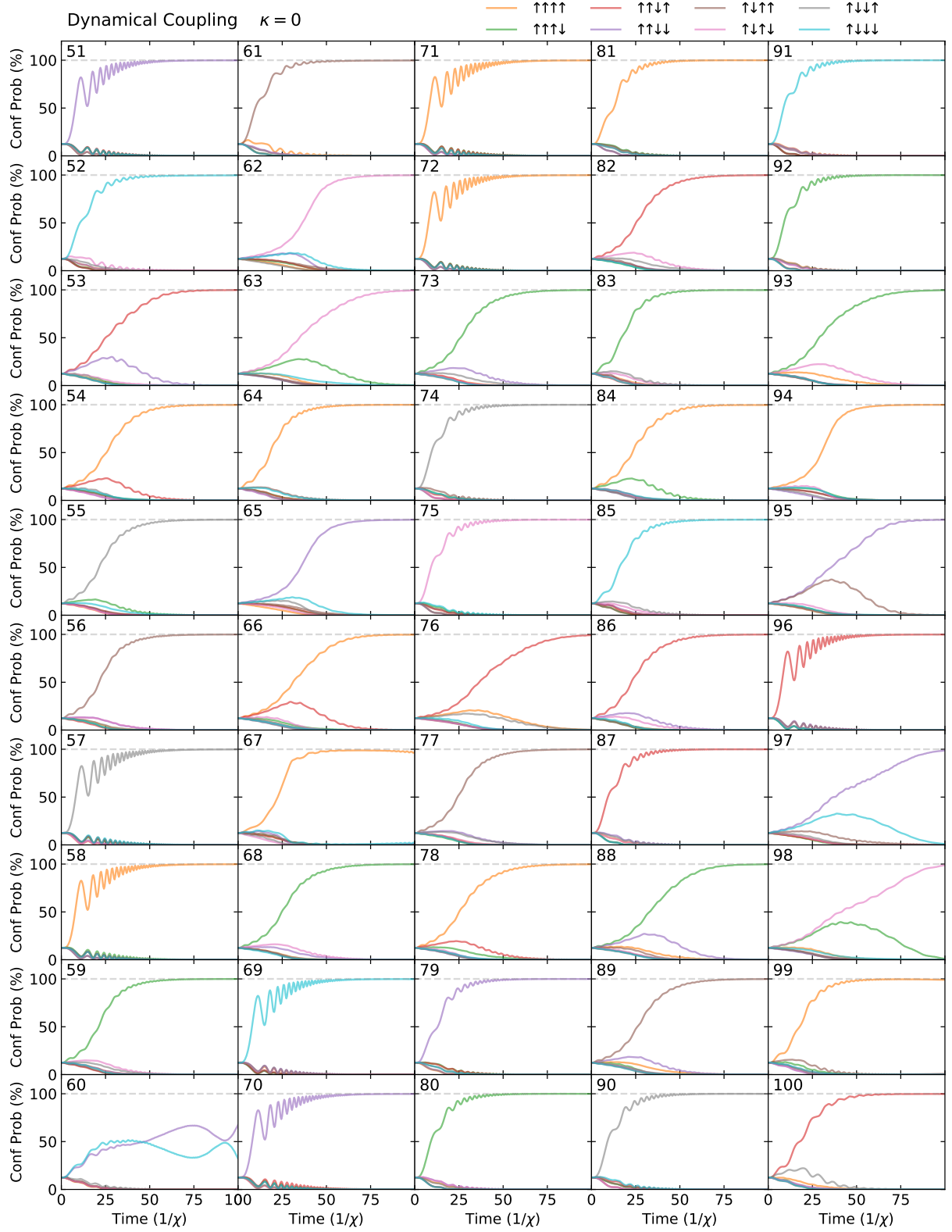
Supplementary Figure 27. Configuration probabilities of statically-coupled system, high-loss case (part 1/2). Evolution of the probabilities ($P_{\sigma}(t)$) of obtaining each of the possible spin configurations in the statically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 17. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



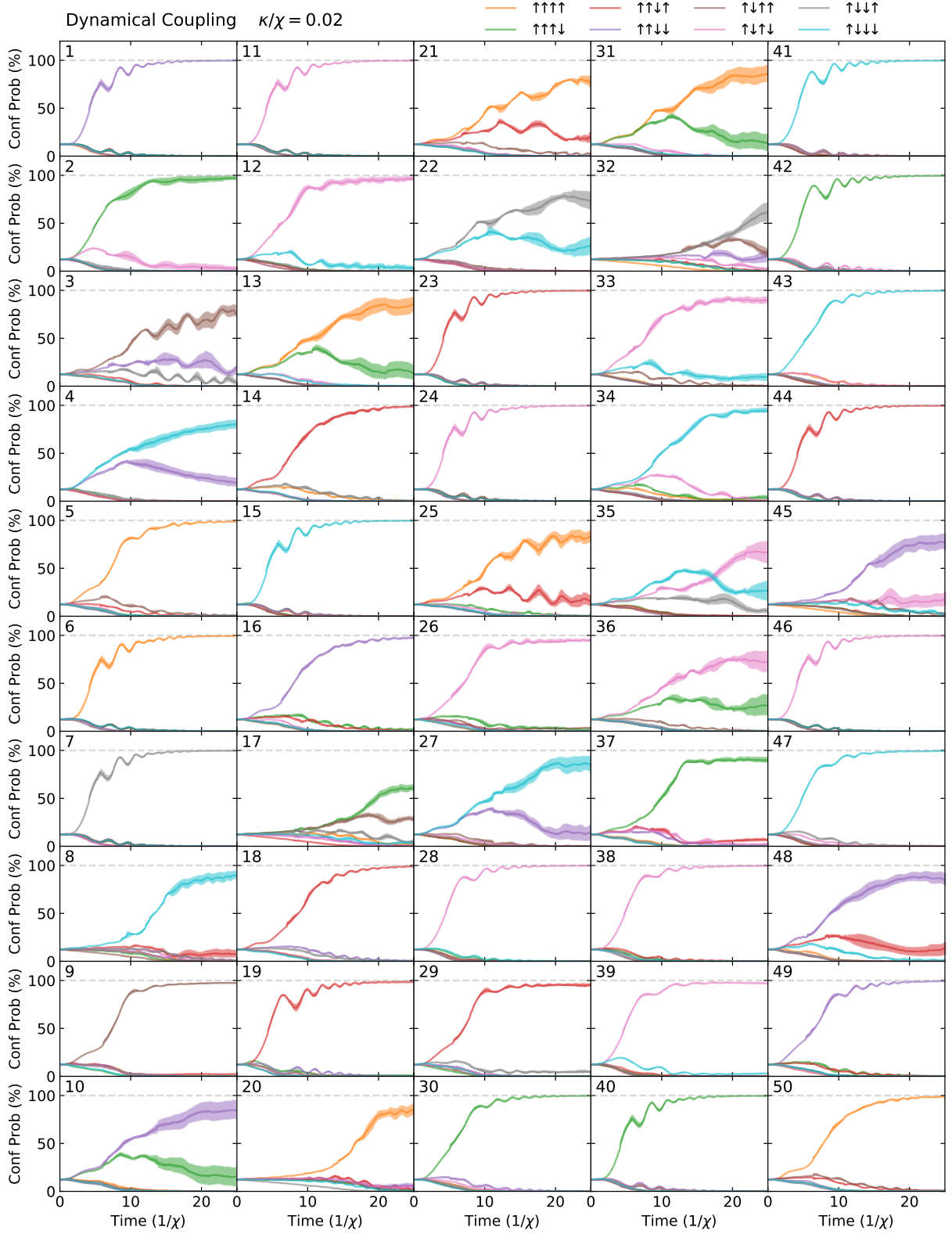
Supplementary Figure 28. Configuration probabilities of statically-coupled system, high-loss case (part 2/2). Evolution of the probabilities ($P_\sigma(t)$) of obtaining each of the possible spin configurations in the statically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 18. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



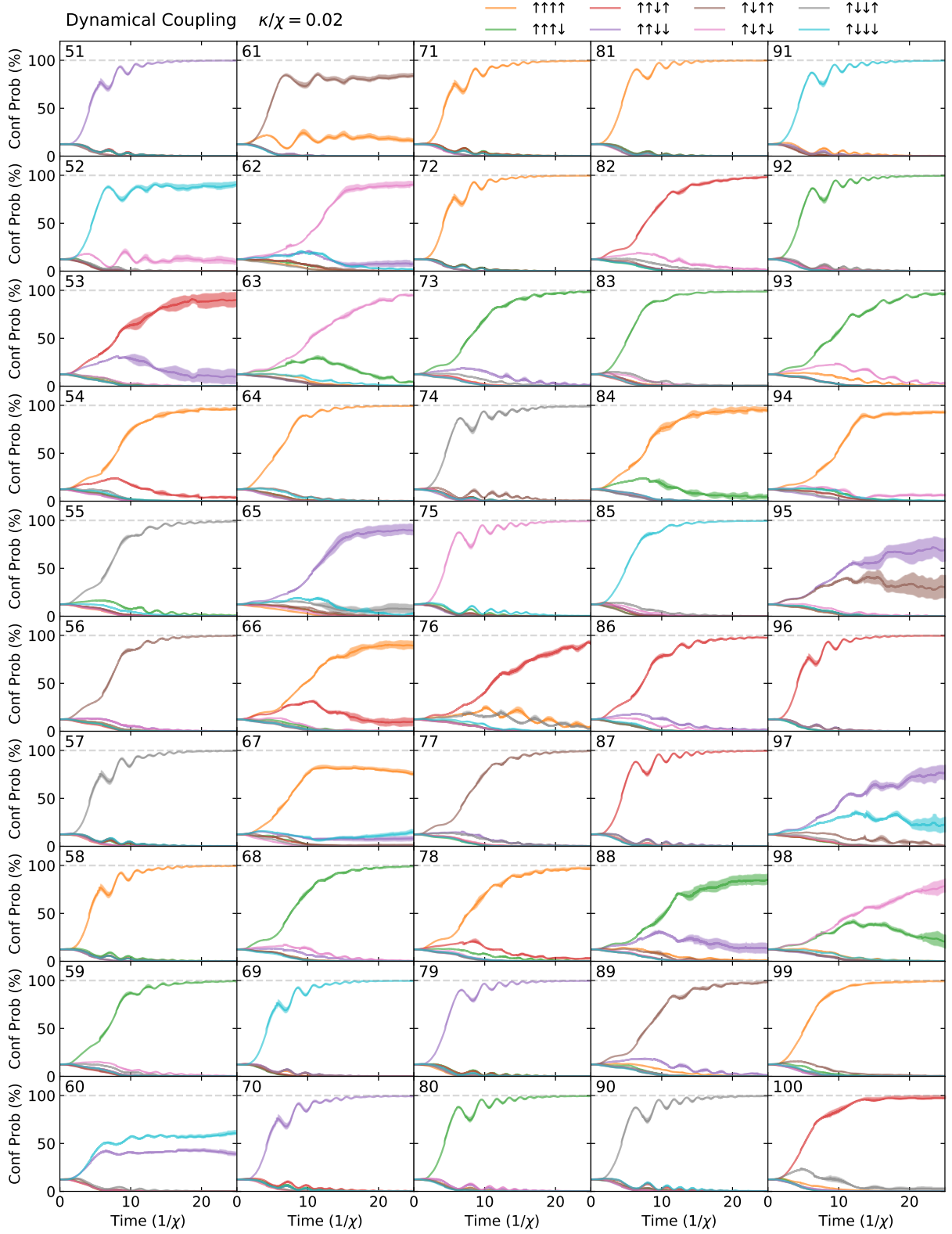
Supplementary Figure 29. Configuration probabilities of dynamically-coupled system, lossless case (part 1/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the dynamically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 17. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology.



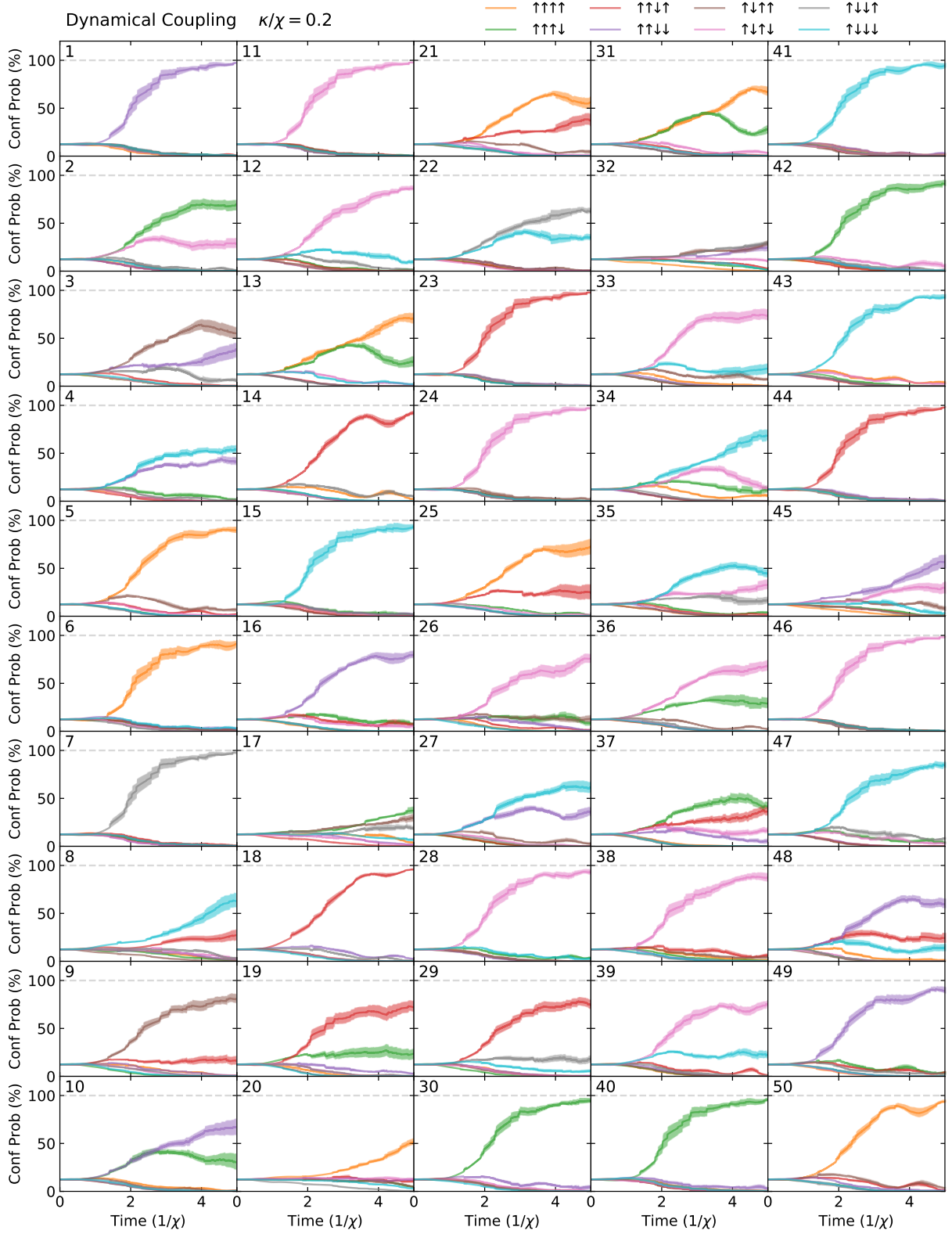
Supplementary Figure 30. Configuration probabilities of dynamically-coupled system, lossless case (part 2/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the dynamically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 18. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology.



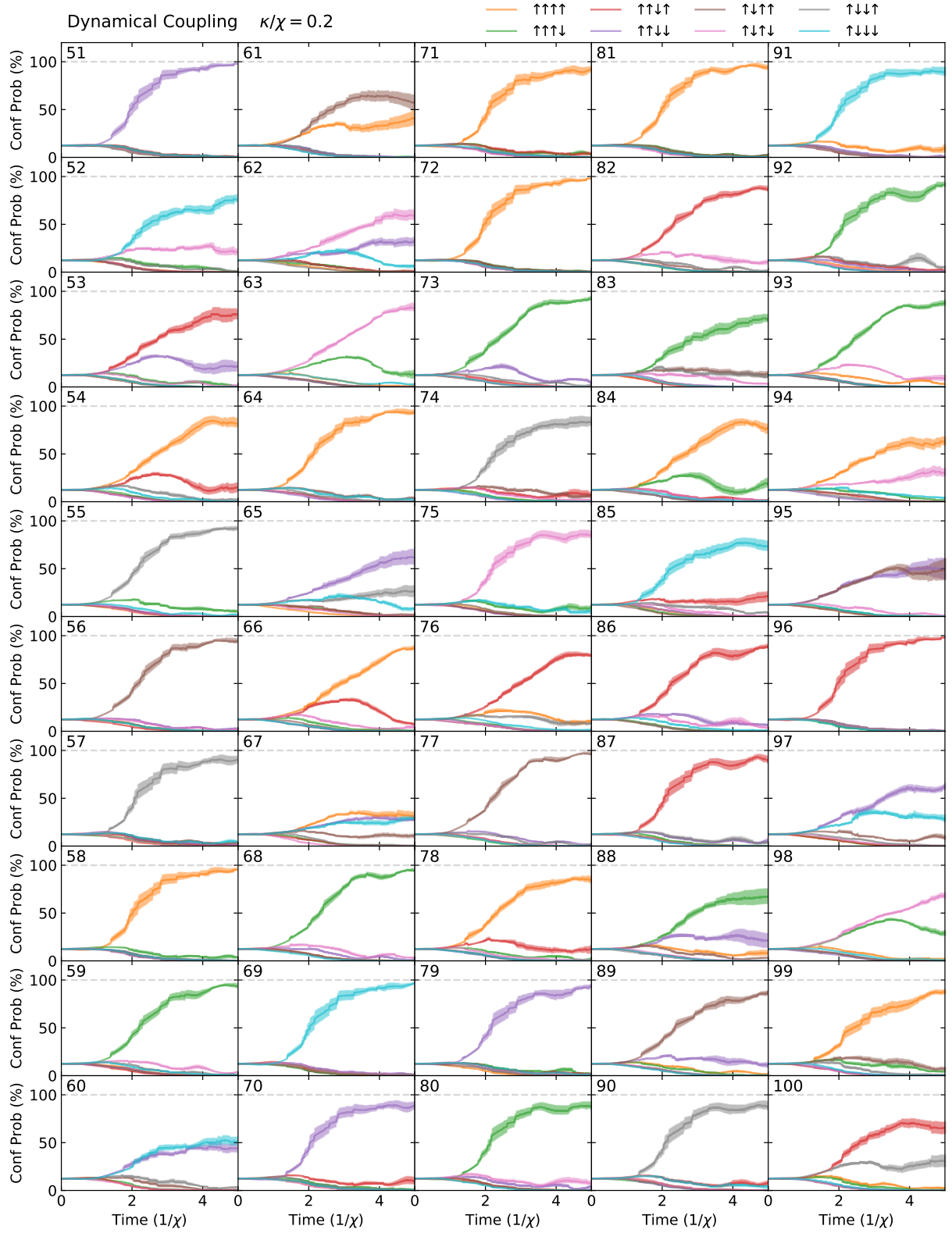
Supplementary Figure 31. Configuration probabilities of dynamically-coupled system, low-loss case (part 1/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the dynamically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 17. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



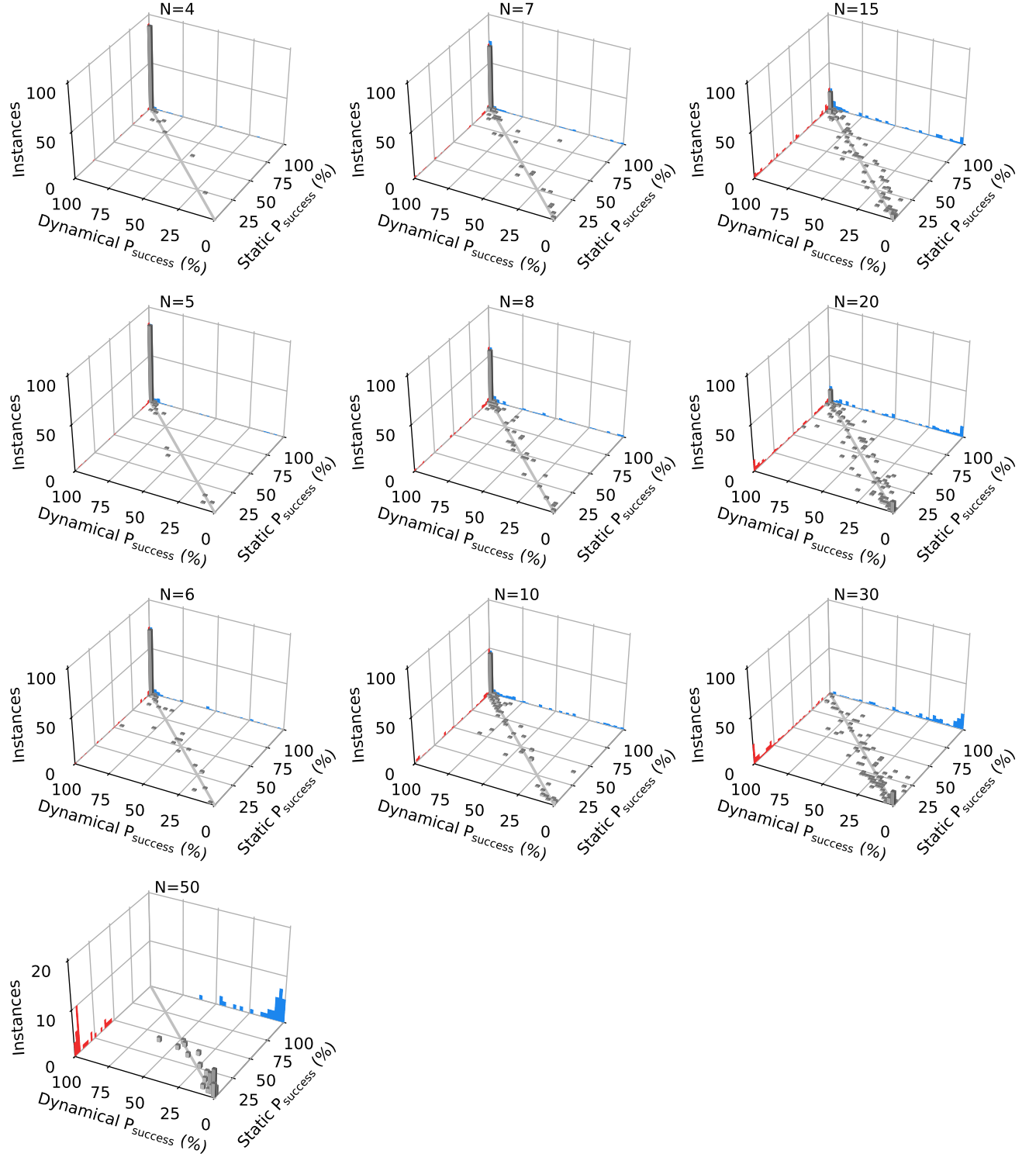
Supplementary Figure 32. Configuration probabilities of dynamically-coupled system, low-loss case (part 2/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the dynamically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 18. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



Supplementary Figure 33. Configuration probabilities of dynamically-coupled system, high-loss case (part 1/2). Evolution of the probabilities $P_{\sigma}(t)$ of obtaining each of the possible spin configurations in the dynamically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 17. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



Supplementary Figure 34. Configuration probabilities of dynamically-coupled system, high-loss case (part 2/2). Evolution of the probabilities $P_\sigma(t)$ of obtaining each of the possible spin configurations in the dynamically-coupled system as a function of the evolution time, for the problem instances depicted inset in the corresponding sub-axis (see upper-left corner) of Sup. Figure 18. The configuration probability is defined according to (93) in Sup. Note 6 A. Note that due to the symmetries of the system, the configuration probability is exactly the same upon flipping every spin, so we take the convention that each curve represents the sum of the probabilities for the two degenerate configurations. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 A for additional details of our quantum simulation methodology. Due to the stochastic nature of the simulation, $N_{\text{traj}} = 10$ trajectories are run for each problem instance; the solid line indicates the mean over the trajectories, while the fill indicates the standard error of the mean, as described in Sup. Note 6 A.



Supplementary Figure 35. Correlation of success probabilities between statically- and dynamically-coupled classical systems. For various problem sizes N , we show the correlation matrix of success probabilities P_{success} (evaluated at the end of the trajectory) between the statically-coupled and dynamically-coupled systems, using the classical EOMs for each respective system. Projected to the sides in blue and red are the corresponding histograms for the dynamically-coupled and statically-coupled systems, respectively. The system parameters are chosen according to the prescription described in Sup. Note 5 A for each Ising problem instance. See Sup. Note 6 B for additional details on our classical simulation methodology.

Supplementary References

- [1] Goto, H. Bifurcation-based adiabatic quantum computation with a nonlinear oscillator network. *Sci. Rep.* **6**, 21686 (2016).
- [2] Bukov, M., D'Alessio, L. & Polkovnikov, A. Universal high-frequency behavior of periodically driven systems: from dynamical stabilization to Floquet engineering. *Adv. Phys.* **64**, 139–226 (2015).
- [3] Mogensen, P. & Riseth, A. Optim: A mathematical optimization package for Julia. *J. Open Source Softw.* **3**, 615 (2018).
- [4] Nigg, S. E., Lörch, N. & Tiwari, R. P. Robust quantum optimizer with full connectivity. *Sci. Adv.* **3**, e1602273 (2017).
- [5] Puri, S., Andersen, C. K., Grimsmo, A. L. & Blais, A. Quantum annealing with all-to-all connected nonlinear oscillators. *Nat. Commun.* **8**, 15785 (2017).
- [6] Farhi, E., Goldstone, J., Gutmann, S. & Sipser, M. Quantum Computation by Adiabatic Evolution. Preprint at <http://arxiv.org/abs/quant-ph/0001106> (2000).
- [7] Goto, H. Universal quantum computation with a nonlinear oscillator network. *Phys. Rev. A* **93**, 050301 (2016).
- [8] Puri, S., Boutin, S. & Blais, A. Engineering the quantum states of light in a Kerr-nonlinear resonator by two-photon driving. *npj Quantum Inf.* **3**, 18 (2017).
- [9] Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge University Press, 2009).
- [10] Frattini, N. E., Sivak, V. V., Lingenfelter, A., Shankar, S. & Devoret, M. H. Optimizing the Nonlinearity and Dissipation of a SNAIL Parametric Amplifier for Dynamic Range. *Phys. Rev. Appl.* **10**, 054020 (2018).
- [11] Koch, J. et al. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
- [12] Girvin, S. M. Circuit QED: superconducting qubits coupled to microwave photons. In *Quantum Machines: Measurement and Control of Engineered Quantum Systems* 113–239 (Oxford Univ. Press, 2014).
- [13] Gardiner, C. & Zoller, P. *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics* (Springer-Verlag Berlin Heidelberg, 2004).
- [14] Song, C. et al. 10-Qubit Entanglement and Parallel Logic Operations with a Superconducting Circuit. *Phys. Rev. Lett* **119**, 180511 (2017).
- [15] Johansson, J., Nation, P. & Nori, F. QuTiP 2: A Python framework for the dynamics of open quantum systems. *Comput. Phys. Commun.* **184**, 1234–1240 (2013).
- [16] Goto, H., Tatsumura, K. & Dixon, A. R. Combinatorial optimization by simulating adiabatic bifurcations in nonlinear Hamiltonian systems. *Sci. Adv.* **5**, eaav2372 (2019).
- [17] Rackauckas, C. & Nie, Q. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *J. Open Res. Softw.* **5**, 15 (2017).
- [18] Sage, J. M., Bolkhovsky, V., Oliver, W. D., Turek, B. & Welander, P. B. Study of loss in superconducting coplanar waveguide resonators. *J. Appl. Phys.* **109**, 063915 (2011).
- [19] Calusine, G. et al. Analysis and mitigation of interface losses in trenched superconducting coplanar waveguide resonators. *Appl. Phys. Lett.* **112**, 062601 (2018).