enough to cover all possible states. We define the two sets of phase-point operators (see Fig. 2 in [40] for a visualization) via

$$r^1(0,0) = (1,1,1), \tag{A2a}$$
$$r^1(0,1) = (-1,-1,1), \tag{A2b}$$
$$r^1(1,0) = (1,-1,-1), \tag{A2c}$$
$$r^1(1,1) = (-1,1,-1), \tag{A2d}$$

and

$$r^2(0,0) = (1,-1,1), \tag{A3a}$$
$$r^2(0,1) = (-1,1,1), \tag{A3b}$$
$$r^2(1,0) = (1,1,-1), \tag{A3c}$$
$$r^2(1,1) = (-1,-1,-1). \tag{A3d}$$

With this choice, we can define the Wigner function of some quantum state $\hat{\rho}$ as

$$w^s(p,q;\hat{\rho}) = \tfrac{1}{2}\operatorname{Tr}\hat{\rho}\hat{A}^s_{(p,q)}, \tag{A4}$$

where $s = 1, 2$ denotes the set of phase-point operators. These are normalized for each set, i.e., $\sum_{p,q} w^s(p,q;\hat{\rho}) = \operatorname{Tr}\hat{\rho} = 1$ independent of s, and thus quasiprobability distributions. In case all values of a $w^s(p,q;\hat{\rho})$ are positive, we can treat it as a probability distribution and sample initial conditions for the truncated Wigner approximation from it. For a single spin, it is always possible to rotate the phase-point operators to render the Wigner functions positive, so we can always sample from either one of the two possible choices of Wigner functions. In fact, it is crucial to employ both choices for sampling to prevent the introduction of spurious correlations (see example below) [33,40]. To be explicit, the complete sampling procedure for a single trajectory first randomly selects one of the phase-space representation and then draws a phase-space vector according to its Wigner function.

We illustrate this prescription using the state $\rho = |{\uparrow}\rangle\langle{\uparrow}|$ as an example. The Wigner functions read as

$$\mathbf{w}^1(\rho) = \begin{pmatrix} w^1(0,0) & w^1(0,1) \\ w^1(1,0) & w^1(1,1) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{pmatrix}, \tag{A5a}$$

$$\mathbf{w}^2(\rho) = \begin{pmatrix} w^2(0,0) & w^2(0,1) \\ w^2(1,0) & w^2(1,1) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{pmatrix}. \tag{A5b}$$

Choosing one these Wigner functions at random and then sampling from it is equivalent to drawing a sample from the set $\{r^1(0,0), r^1(0,1), r^2(0,0), r^2(0,1)\}$. In turn, this just means we need to set the z component to 1 and choose x and y independently from ± 1. We remark that this prescription reproduces all moments of the spin operators $\langle(\hat{\sigma}_{x,y,z})^k\rangle$ in contrast to the Gaussian approximation which reproduces means and covariances only. Additionally, we remark that every possible phase point of $\mathbf{w}^1$ ($\mathbf{w}^2$) has the x and y components aligned (antialigned), which is the spurious correlation mentioned earlier. By using both Wigner functions, we avoid artifacts caused by this, making the simulation more accurate.

2. Generalization to clusters of spins

The prescription, we just outlined, readily generalizes to clusters of spins by taking tensor products of the phase-point operators. Consider a cluster of n spins: The joint Hilbert space is now SU(D), where $D = 2^n$, given from the tensor product of Hilbert spaces of the single spins. In the following, we essentially repeat the construction from before applied to the cluster's Hilbert space and exploit its product structure. We denote the operator basis of a cluster of n spins by $\mathbf{X}_n$, which can be constructed recursively by

$$\hat{\mathbf{X}}_1 \equiv \hat{\sigma}, \tag{A6a}$$
$$[\hat{\mathbf{X}}_n]_i = \begin{cases} \hat{\sigma}_i \otimes \mathbb{1}, & i \in \{1,2,3\} \\ \mathbb{1} \otimes \hat{\sigma}_{i-3}, & i \in \{4,5,6\} \\ [\hat{\mathbf{X}}_1 \otimes \hat{\mathbf{X}}_{n-1}]_{i-6}, & \text{else} \end{cases} \tag{A6b}$$

where $[\cdot]_i$ denotes the ith component of the vector.

In much the same way, we can construct the phase-point vectors. However, we need to consider that we have two possible choices for each spin to make, so there are a total of 2^n sets of phase-point operators. Using $\mathbf{s} \in \{1,2\}^n$, we can construct the phase-point vectors corresponding to the operator basis defined above as

$$\begin{aligned} \mathbf{r}^{\mathbf{s}}_n(\mathbf{p},\mathbf{q}) = {} & \mathbf{r}^{s_1}(p_1,q_1) \\ & \oplus \mathbf{r}^{\tilde{\mathbf{s}}}_{n-1}(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \\ & \oplus \left[\mathbf{r}^{s_1}(p_1,q_1) \otimes \mathbf{r}^{\tilde{\mathbf{s}}}_{n-1}(\tilde{\mathbf{p}},\tilde{\mathbf{q}})\right], \end{aligned} \tag{A7}$$

where the vectors with tilde ($\tilde{\mathbf{s}}$, $\tilde{\mathbf{p}}$, and $\tilde{\mathbf{q}}$) are the same as the bare vectors without the first element, e.g., $\tilde{\mathbf{s}} = (s_2, \ldots, s_n)$. From these building blocks, we can define the Wigner functions of the cluster as

$$w^{\mathbf{s}}(\mathbf{p},\mathbf{q};\hat{\rho}) = \frac{1}{2^n}\operatorname{Tr}\hat{\rho}\hat{A}^{\mathbf{s}}_{(\mathbf{p},\mathbf{q})} \tag{A8}$$
$$= \frac{1}{4^n}\operatorname{Tr}\hat{\rho}\left(\mathbb{1} + \mathbf{r}^{\mathbf{s}}_n(\mathbf{p},\mathbf{q})\cdot\hat{\mathbf{X}}_n\right). \tag{A9}$$

As can be checked easily via induction, this definition gives us a normalized Wigner function for every choice of $\mathbf{s}$.

Another short calculation shows that if the quantum state $\hat{\rho}$ factorizes between the spins, i.e., $\rho = \bigotimes_{i\leqslant n}\hat{\rho}_i$, then the Wigner function factorizes as well:

$$w^{\mathbf{s}}\left(\mathbf{p},\mathbf{q};\bigotimes_{i\leqslant n}\hat{\rho}_i\right) = \prod_{i\leqslant n} w^{s_i}(p_i,q_i;\hat{\rho}_i). \tag{A10}$$

This allows for efficient sampling.

To derive rules for sampling initial states, conceptually one needs to choose a random set of phase-point representations, i.e., draw $\mathbf{s}$ randomly, and then choose a phase-space vector $\mathbf{r}^{\mathbf{s}}_n(\mathbf{p},\mathbf{q})$ with a probability determined by the corresponding Wigner function $w^{\mathbf{s}}(\mathbf{p},\mathbf{q})$. In case of a product initial state, this prescription simplifies dramatically because we choose the phase-space vector of each spin independently and compute the initial value of correlators by products [see Eq. (A8)].