Block Encoding

Use a block encoding as a scaled matrix interface inside a larger reversible operation.

Research 22 min lesson Dr. Charith Jayasekara

By the end, you'll be able to

- State block scale, ancillas and approximation error.
- Verify a small unitary dilation.
- Calculate input-dependent branch probabilities.

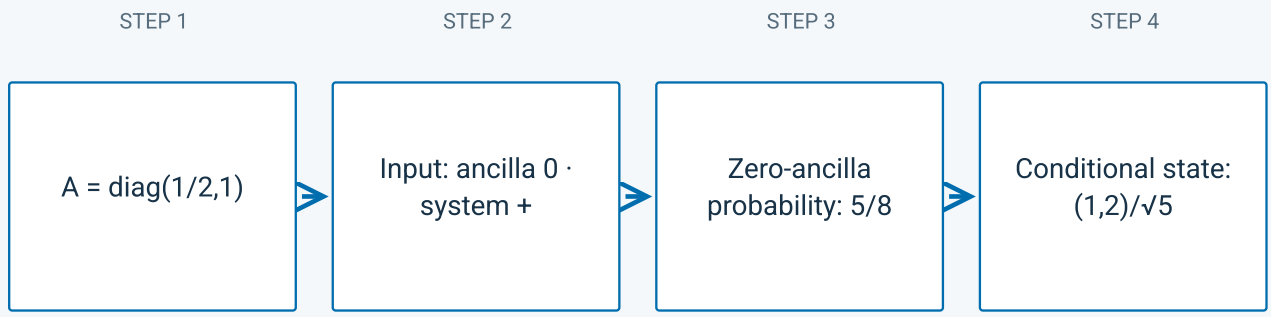
Hide a useful matrix inside a larger unitary

A block encoding places a scaled matrix inside a designated block of a unitary acting on a larger system. The extra ancillas supply the room needed to preserve reversibility even when the matrix itself is not unitary. This is an interface for matrix algorithms: once useful access is available, other constructions can transform its spectrum or singular values.

Prerequisites: linear algebra, tensor products and LCU. The QSVT primary paper develops block encodings as part of a unified matrix-algorithm framework. Gilyén and colleagues.

WORKED DIAGRAM · A27

A diagonal block encoding



The complementary ancilla branch preserves total norm. The selected block need not be a deterministic unitary on the system.

The setup includes scale and error

Designate ancillas that begin and end in zero. Projecting the full unitary onto those ancilla zeros yields A divided by alpha, up to a stated approximation error. Alpha must be large enough that the encoded block has operator norm at

most one. The task description should also state ancilla count, implementation cost and which controlled or inverse accesses are available.

An arbitrary large matrix can have a mathematically valid unitary dilation without having an efficient implementation. The distinction between existence and efficient access is fundamental. A theorem counting calls to a block encoding does not include the construction cost unless that cost is explicitly supplied.

The scale and approximation error belong in the mathematical definition of the interface.

$$B = (\langle 0^a| \otimes I)U(|0^a\rangle \otimes I), \qquad \|A - \alpha B\| \le \epsilon, \qquad P(0^a \mid \psi) = \||B|\psi\rangle\|^2.$$

U is the full unitary, a the number of ancillas, α the scale and ϵ an operator-norm error measured in the units of A. An exact encoding has $B = A/\alpha$. In the worked example $a = 1, \alpha = 1, \epsilon = 0$, and input plus gives probability $5/8$.

Worked example: a diagonal setupion

Let A be the diagonal matrix with entries one half and one. Its norm is one, so choose alpha equal to one. Let S be the diagonal matrix with entries square root of three over two and zero. Because A and S commute and their squares sum to the identity, the larger block matrix

$$U = \begin{pmatrix} A & S \\ S & -A \end{pmatrix}$$

is unitary. Multiplying it by its adjoint gives identity in the diagonal blocks and zero in the off-diagonal blocks. The upper-left block is exactly A. The block order corresponds to the ancilla label, while each block acts on the original one-qubit system.

Apply U to ancilla zero and system plus. The zero-ancilla system component is the column one half, one divided by square root of two. Its squared norm is five eighths. Conditional on that ancilla result, the system becomes the column one, two divided by square root of five, with Z probabilities one fifth and four fifths.

The other ancilla branch retains the remaining norm so the full operation stays unitary. This is why a contraction can be applied conditionally without violating norm preservation. The complete state is normalised even though its selected block is not.

A block is an interface, not always a measurement instruction

Postselecting ancilla zero is one way to use the block, but many algorithms keep ancillas coherent and use reflections or polynomial sequences. Measuring and discarding failure branches at every intermediate step can change both the algorithm and its complexity. Follow the intended construction rather than assuming every block call should be measured immediately.

For a nonunitary A, success probability generally depends on the input through the norm of A acting on it. Amplitude amplification may improve a postselected workflow if its coherent prerequisites are available. Report those reflections, inverse calls and input-state preparations. They are resources, not a free conversion from a matrix into a deterministic gate.

How useful block encodings arise

LCU gives one route: prepare a coefficient index, select a unitary term and undo the index preparation. Sparse-access constructions give another route when matrix entries and locations can be queried efficiently. Structured arithmetic, quantum walks and data structures can provide other representations. Each route makes different assumptions about how the matrix is supplied.

For a classical data matrix, loading costs may dominate. If the matrix changes frequently, a costly preprocessing structure may need rebuilding. If it is fixed and reused, amortisation may be relevant. State the application workload before deciding which costs can be shared across calls.

Compose and test the interface

Check the desired block on every basis input for small examples. Compare complex amplitudes, not just probabilities. Verify the full unitary norm and the declared block scaling. Then test a superposition with imaginary entries to catch phase or conjugation errors that real basis tests can conceal in a numeric implementation.

Composition requires care. Products of block encodings often need separate ancillas and have multiplied scale factors. Sums can use an index construction with their own normalisation. Approximation errors accumulate through those operations. A diagram showing one box per mathematical matrix can hide a growing ancilla register and large scaling factors.

A scale factor changes the success calculation

Encode twice the worked diagonal matrix using scale alpha equal to two. The encoded block is unchanged, so the ancilla-zero probabilities on each input are unchanged. A report quoting the unscaled matrix without alpha would nevertheless misidentify which linear operation the block represents. The scale is part of the interface's mathematical meaning.

For input plus, the conditional state remains proportional to one, two, while its probability follows the encoded contraction rather than the norm of the unscaled matrix alone. Separating target matrix, block scale and conditional normalisation prevents a common error in matrix-algorithm demonstrations.

When composing several such interfaces, track every scale symbol through the algebra. A product may encode the desired product divided by the product of scales. Ignoring those factors can turn a correct-looking conditional state into an unrealistic claim of deterministic success or understated amplification cost.

Limits and connections

Block encoding is a representation primitive. It does not by itself invert A, prepare an eigenvector or return matrix entries. Qubitization and QSVT use the interface for further transformations. Their costs depend on scale, approximation degree, spectral promises and output extraction. A large matrix dimension alone is not enough to establish an advantage.

The toy dilation is deliberately explicit and inefficient as a general loading prescription. It proves the algebraic idea on a small example. For a scalable algorithm, replace “write down this unitary” with an actual circuit or efficient oracle construction and count the gates required.

Exercises and worked answers

Exercise 1. Why is alpha equal to one invalid for A equal to twice the identity?

Check the worked answer

A block of a unitary cannot have norm greater than one. Scale by at least two, so the encoded block is identity or another contraction. The larger scale affects subsequent algorithmic costs.

Exercise 2. For the worked diagonal A, what is the success probability on input zero and on input one?

Check the worked answer

It is one quarter on zero and one on one. The norm of A times the input determines the selected branch probability. The full larger unitary remains norm-preserving in both cases.

Check your understanding

PAUSE & THINK

For the worked A, what is the ancilla-zero probability on input |0⟩?

- [ ] 1/4
- [ ] 1/2
- [ ] 1

Answer: 1/4. A|0⟩ has amplitude 1/2, so its squared norm is 1/4. The rest lies in the complementary branch.

Sources

- Gilyén and colleagues