

## Arquiteturas Modulares e Redes Quânticas: Um estudo da atualidade por meio de revisão bibliográfica

Modular Architectures and Quantum Networks: A current study through bibliographic review

Arquitecturas Modulares y Redes Cuánticas: Un estudio actual a través de revisión bibliográfica

Recebido: 15/11/2025 | Revisado: 01/12/2025 | Aceitado: 02/12/2025 | Publicado: 04/12/2025

**Tarcisio De Souza Peres**

ORCID: <https://orcid.org/0009-0009-2801-0896>

Universidade Paulista, Brasil

E-mail: [Dr.tarcisio@usp.br](mailto:Dr.tarcisio@usp.br)

**Cláudio Boghi**

ORCID: <https://orcid.org/0000-0002-7974-6416>

Universidade Paulista, Brasil

E-mail: [claudio.boghi@docente.unip.br](mailto:claudio.boghi@docente.unip.br)

### Resumo

As limitações físicas e de engenharia dos dispositivos quânticos atuais impulsionam o desenvolvimento de arquiteturas modulares e redes quânticas para viabilizar computadores quânticos em grande escala. Este artigo tem o objetivo de apresentar uma revisão bibliográfica sobre arquiteturas modulares em computação quântica. No estudo se busca-se apresentar e discutir as principais plataformas de hardware - incluindo qubits supercondutores, íons aprisionados, átomos neutros e qubits fotônicos - e analisando como módulos quânticos podem ser interconectados os métodos de entrelaçamento distribuído, repetidores quânticos e teleporte de portas lógicas, bem como modelos de custo e latência para operações quânticas distribuídas. Aspectos de escalabilidade e desafios práticos de implementação física são examinados à luz de propostas e estudos de caso de arquiteturas modulares reais, como as abordagens da IonQ e Honeywell (Quantinuum) com íons aprisionados e da Xanadu com fotônica. Por fim, comparamos as abordagens de diferentes plataformas e discutimos caminhos para superar obstáculos técnicos rumo a computadores quânticos modularmente escaláveis e tolerantes a falhas.

**Palavras-chave:** Computação quântica; Hardware; Arquitetura computacional; Qubits.

### Abstract

The physical and engineering limitations of current quantum devices are driving the development of modular architectures and quantum networks to enable large-scale quantum computers. This article aims to present a literature review on modular architectures in quantum computing. The study aims to present and discuss the main hardware platforms—including superconducting qubits, trapped ions, neutral atoms, and photonic qubits—and analyze how quantum modules can be interconnected with distributed entanglement methods, quantum repeaters, and logic gate teleportation, as well as cost and latency models for distributed quantum operations. Scalability aspects and practical physical implementation challenges are examined in light of proposals and case studies of real modular architectures, such as the IonQ and Honeywell (Quantinuum) approaches with trapped ions and Xanadu with photonics. Finally, we compare the approaches of different platforms and discuss ways to overcome technical obstacles toward modularly scalable and fault-tolerant quantum computers.

**Keywords:** Quantum computing; Hardware; Computational architecture; Qubits.

### Resumen

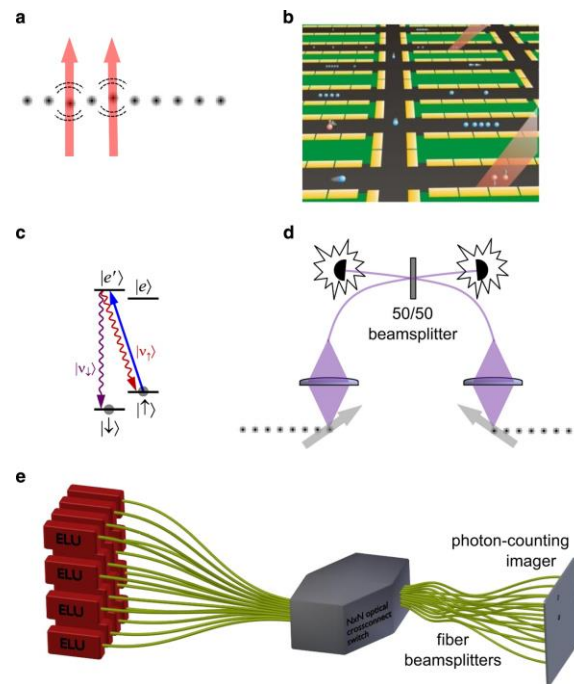
Las limitaciones físicas y de ingeniería de los dispositivos cuánticos actuales impulsan el desarrollo de arquitecturas modulares y redes cuánticas para la creación de computadoras cuánticas a gran escala. Este artículo presenta una revisión bibliográfica sobre arquitecturas modulares en computación cuántica. El estudio busca presentar y discutir las principales plataformas de hardware —incluyendo cúbits superconductores, iones atrapados, átomos neutros y cúbits fotónicos— y analizar cómo los módulos cuánticos pueden interconectarse con métodos de entrelazamiento distribuido, repetidores cuánticos y teletransportación de puertas lógicas, así como modelos de costo y latencia para operaciones cuánticas distribuidas. Se examinan los aspectos de escalabilidad y los desafíos prácticos de la implementación física a la luz de propuestas y casos prácticos de arquitecturas modulares reales, como los enfoques IonQ y Honeywell (Quantinuum) con iones atrapados y Xanadu con fotónica. Finalmente, comparamos enfoques en diferentes plataformas y analizamos maneras de superar los obstáculos técnicos para lograr computadoras cuánticas modularmente escalables y tolerantes a fallos.

**Palabras clave:** Computación cuántica; Hardware; Arquitectura computacional; Qubits.

## 1. Introdução

À medida que a computação quântica evolui de protótipos de laboratório para máquinas cada vez maiores, surgem desafios substanciais de escalabilidade. Qubits adicionais acarretam complexidade de controle, aumento de interações espúrias e dificuldade de manter coerência. Assim como na computação clássica o paradigma multi-core mitigou limitações de clock e dissipação, na computação quântica desponta a ideia de arquiteturas modulares. Nessa abordagem, em vez de um único processador monolítico de centenas ou milhares de qubits, constrói-se um sistema a partir de múltiplos módulos quânticos menores, interligados por uma rede quântica que distribui entrelaçamento entre eles (Monroe, 2014; Nickerson, Fitzsimons & Benjamin, 2014). Cada módulo pode executar operações locais de forma rápida e coesa, enquanto operações entre módulos são viabilizadas pelo compartilhamento de pares de qubits entrelaçados por meio de interconexões fotônicas, canais de micro-ondas ou outras interfaces quânticas apropriadas (Monroe, 2014). A Figura 1, abaixo, ilustra esse conceito no contexto de íons aprisionados em múltiplos módulos interconectados.

**Figura 1:** Exemplo de arquitetura modular proposta para computação quântica com íons aprisionados.



Fonte: Brown (2016).

Como afirmam Brown (2016), verifica-se pela Figura 1 que cada módulo (ELU) contém 50–100 íons com interações locais via modos coletivos (fonons), enquanto entre módulos o entrelaçamento é distribuído por uma rede fotônica com comutadores ópticos e interferômetros que permitem conexões arbitrárias entre qubits de módulos distintos.

Essa arquitetura modular visa contornar limitações inerentes dos dispositivos atuais. Por exemplo, em dispositivos supercondutores monolíticos, a adição de muitos qubits torna complexa a interconexão on-chip e aumenta erros correlacionados; em íons aprisionados, cristais muito longos degradam a velocidade de porta e a fidelidade devido a modos vibracionais numerosos e ruído ambiental; já sistemas fotônicos sofrem com perdas óticas crescentes conforme se expande a dimensão do circuito. Ao segmentar o processador quântico em módulos menores (cada qual mais fácil de controlar com alta fidelidade) e conectá-los por enlaces quânticos, espera-se obter escalabilidade fault-tolerant (tolerante a falhas). Monroe et al. destacam que a construção prática de um computador quântico de grande porte provavelmente exigirá a justaposição de

diferentes tipos de sistemas quânticos – por exemplo, qubits de memória interconectados por mediadores fotônicos – formando uma hierarquia de interações locais e remotas capaz de escalar para um número muito grande de qubits (Monroe, 2014). Em 2014, esses autores propuseram uma arquitetura modular com íons aprisionados que ilustra essa filosofia: dentro de cada módulo, portas de dois qubits são realizadas por interações naturais (e.g. Coulomb) entre íons vizinhos, enquanto entre módulos distantes as portas quânticas são efetuadas via entrelaçamento distribuído por interfaces fotônicas entre qubits de comunicação (Ibid, 2014). Essa arquitetura mostrou-se, em princípio, escalável e compatível com protocolos de correção de erros quânticos, apresentando potencial para execução tolerante a falhas de algoritmos em um grande número de qubits (Monroe, 2014).

Nos últimos anos, demonstrações experimentais validaram elementos fundamentais das arquiteturas modulares. Em 2014, Nickerson et al. demonstraram, via simulações, que células quânticas de 5–50 qubits conectadas por enlaces fotônicos com perdas de até 98% ainda podem operar coletivamente em uma rede quântica tolerante a falhas, mediante protocolos de purificação de entrelaçamento e códigos de correção de erros de superfície (Nickerson, Fitzsimons & Benjamin, 2014). Embora enlaces com perdas elevadas limitem a taxa de entrelaçamento remoto para dezenas ou centenas de hertz (tipicamente ~100 Hz para íons aprisionados), esses vínculos podem ser utilizados de forma eficiente por meio de diversos enlaces paralelos. A disponibilidade de vários pares entrelaçados em paralelo permite que a taxa efetiva de operações remotas seja adequada para alguns algoritmos. (Brown, Kim & Monroe, 2016). Em 2021, Gold et al. demonstraram o acoplamento determinístico entre quatro chips supercondutores separados, obtendo fidelidades de 99,1 % (iSWAP) e 98,3 % (CZ), valores comparáveis às portas internas em um único chip. Esses avanços indicam que módulos quânticos independentes podem ser conectados mantendo alto desempenho, apoiando a visão de computadores quânticos modularmente construídos.

Este artigo tem o objetivo de apresentar uma revisão bibliográfica sobre arquiteturas modulares em computação quântica.

O presente estudo está organizado da seguinte forma: inicialmente revisamos as plataformas físicas relevantes para arquiteturas modulares (Seção 2), discutindo as características de qubits supercondutores, íons aprisionados, átomos neutros e fotônicos. Em seguida, abordamos os meios de conectividade entre módulos e suas limitações físicas (Seção 3). A Seção 4 explora conceitos de redes quânticas, incluindo distribuição de entrelaçamento, repetidores quânticos para longas distâncias e o teleporte de portas lógicas entre nós. Na Seção 5, examinamos modelos de custo e latência de operações distribuídas, fundamentais para avaliar o desempenho de arquiteturas modulares. A Seção 6 trata da escalabilidade e dos desafios práticos de implementação física dessas arquiteturas. Apresentamos então estudos de caso de arquiteturas modulares reais ou propostas (Seção 7), destacando as abordagens da IonQ, Honeywell/Quantinuum e Xanadu. Na Seção 8, comparamos as diferentes abordagens, discutindo vantagens relativas e obstáculos. Finalmente, concluímos resumindo os progressos atuais e perspectivas futuras para arquiteturas modulares e redes quânticas na computação quântica.

## **2. Metodologia**

Realizou-se uma pesquisa documental de fonte indireta do tipo revisão bibliográfica não-sistemática (Snyder, 2019) de natureza qualitativa (Pereira et al., 2018) e, do tipo específico de revisão narrativa da literatura (Rother, 2007).

## **3. Resultados e Discussão**

### **3.1 Plataformas de Hardware Quântico para Arquiteturas Modulares**

Um espectro diverso de tecnologias de qubits está sendo explorado para computação quântica, cada qual com vantagens e desafios específicos no contexto de arquiteturas modulares. Nesta seção, discutimos brevemente quatro plataformas de

destaque – supercondutores, íons aprisionados, átomos neutros e fótons – e como cada uma pode ser empregada em arquiteturas modulares e redes quânticas.

### 3.1.1 Qubits Supercondutores

Qubits supercondutores, como os do tipo transmon, são circuitos elétricos que operam em temperaturas criogênicas (~10–20 mK) e têm se destacado por suas velocidades de operação extremamente altas. Portas lógicas de um e dois qubits podem ser realizadas em dezenas de nanossegundos, com altas fidelidades em dispositivos de última geração. Em processadores supercondutores convencionais, os qubits são integrados em um único chip ou em módulos multi-chip fortemente acoplados, arranjados tipicamente em malhas bidimensionais com conectividade de vizinhos próximos. Essa arquitetura monolítica enfrenta, contudo, obstáculos de escalabilidade: o roteamento de linhas de controle e leitura para centenas de qubits em um único chip é complexo, e a fabricação de um chip grande sem defeitos torna-se cada vez mais difícil conforme o número de qubits aumenta. Além disso, eventos de radiação podem induzir falhas correlacionadas simultaneamente em vários qubits de um único chip criogênico.

Arquiteturas modulares propõem dividir um grande processador supercondutor em múltiplos chips interconectados, mitigando esses problemas. Em 2021, Gold et al. demonstraram um dispositivo supercondutor modular composto de quatro chips de oito qubits cada, acoplados deterministicamente via um substrato de interconexão comum. Esse sistema multi-chip realizou portas entre qubits de chips distintos (separados fisicamente) com fidelidade de ~99%, virtualmente igual às portas intra-chip (Gold. A., et al, 2021). O arranjo utilizou flip-chip bonding e acopladores capacitivos entre dies para prover interações entre qubits de chips adjacentes (Chen, Y. et al., 2021; Awschalom, D. et al., 2021). Esse resultado valida que módulos supercondutores podem ser diretamente conectados com baixa latência e alto desempenho, criando efetivamente um processador unificado maior (Gold. A., et al, 2021).

Para interconectar módulos supercondutores separados por distâncias maiores (por exemplo, em refrigeradores distintos), exploram-se dois enfoques principais: linhas de micro-ondas criogênicas e conversão eletro-óptica. No primeiro caso, cabos coaxiais ou guias de onda supercondutores podem transmitir estados quânticos (fótons de micro-ondas) entre chips ou criostatos próximos. Experimentos já demonstraram a troca direta de informação quântica entre qubits supercondutores através de linhas de micro-ondas, viabilizando entrelaçamento de qubits em módulos separados por até alguns metros (Niu, Jingjing et al., 2023). Contudo, perdas e ruídos térmicos limitam o alcance prático no domínio de micro-ondas – por isso, conexões de longa distância podem exigir converter qubits supercondutores (fótons de micro-ondas) em fótons ópticos telecom, que sofrem perdas muito menores em fibras ópticas. Esse processo de conversão quântica ainda é objeto de intensa pesquisa.

Importante notar que, apesar de possíveis enlaces remotos, os qubits supercondutores ainda terão suas operações locais muito mais rápidas e confiáveis que quaisquer operações entre módulos. Protocolos de entrelaçamento probabilístico (heráldico) em outras plataformas tipicamente atingem taxas de até centenas de hertz (Matsukevich, D. N. et al., 2006), enquanto portas locais supercondutoras operam a dezenas de megahertz com fidelidades próximas de 99,9% (Pino, J. M. et al., 2021). Assim, um desafio nas arquiteturas modulares supercondutoras é que as conexões remotas dificilmente igualarão o desempenho das interações locais. Por isso, a tendência é integrar módulos o mais proximamente possível, como no experimento multi-chip citado – mantendo módulos a distâncias nanométricas ou milimétricas – e usar enlaces realmente longos apenas quando necessário (por exemplo, para interligar diversos mainframes quânticos em um cluster). Ainda assim, a modularidade traz benefícios concretos: isolamento entre módulos reduz interferências e erros correlacionados, e fabricar vários chips menores em vez de um grande melhora o rendimento e permite selecionar os melhores chips para compor o

sistema final (Gonzales, Alvin et al., 2023). Masterizar tecnologias de integração 3D e packaging de múltiplos chips supercondutores é, portanto, uma meta crucial para viabilizar processadores escaláveis nessa plataforma (Gonzales, Alvin et al., 2023).

### 3.1.2 Íons Aprisionados

Qubits de íons aprisionados – átomos carregados confinados em armadilhas eletromagnéticas a vácuo ultra alto – representam uma das plataformas mais maduras em termos de fidelidade e coerência. Estados internos eletrônicos ou de spin nuclear dos íons (como  $^{171}\text{Yb}^+$  ou  $^{40}\text{Ca}^+$ ) formam qubits quase perfeitamente idênticos, que podem ser manipulados por lasers ou micro-ondas com fidelidades de porta de dois qubits superiores a 99,9% em experimentos recentes. Uma característica notável dos íons é que todos os qubits dentro de um mesmo cristal aprisionado compartilham modos coletivos de movimento, fornecendo um canal de interação global: em princípio, qualquer par de íons em uma cadeia linear pode ser acoplado via modos vibracionais mediadores. Isso confere conectividade total em pequenos registros de íons, em contraste com as vizinhanças limitadas em chips supercondutores. À medida que se aumenta a cadeia de íons, surgem modos vibracionais adicionais que tornam as portas de dois qubits mais lentas e suscetíveis a ruído; experimentos práticos apontam que registros lineares confiáveis estão limitados a algumas dezenas de íons. A literatura discute que cadeias maiores exigem segmentação ou transporte de íons (QCCD) para manter alta fidelidade; modos vibracionais adicionais introduzem acoplamentos espúrios e crosstalk; e erros de aquecimento e decoerência motional acumulam-se mais significativamente em escalas de tempo maiores (Brown, Kim & Monroe, 2016). Na prática, estima-se que um único cristal linear consegue operar de forma confiável com algumas dezenas de íons (10–50) antes de esses efeitos se tornarem problemáticos (Brown, Kim & Monroe, 2016).

Duas abordagens complementares estão sendo seguidas para escalar sistemas de íons aprisionados além dessa faixa. A primeira é a arquitetura QCCD (Quantum Charge-Coupled Device), que propõe dividir um conjunto grande de íons em múltiplas zonas de aprisionamento dentro de um mesmo chip de armadilha, e transporte ativo de íons individuais ou pequenos grupos entre essas zonas conforme necessário para realizar operações (Brown, Kim & Monroe, 2016). Essencialmente, em vez de manter todos os  $N$  íons em interação direta simultânea, o QCCD desloca dinamicamente íons portadores de informação através de uma “rede de trilhos” na armadilha para efetuar interações moduladas entre diferentes regiões. Essa ideia, concebida nos anos 2000, vem sendo refinada com demonstrações de transporte adiabático e rápido de íons através de junções complexas sem perda de coerência. A Honeywell (hoje Quantinuum) implementou um sistema do tipo QCCD, atingindo até 10 qubits logicamente totalmente conectados via movimentação, e demonstrando fidelidades de porta controladas acima de 99% mesmo com a complexidade extra do transporte. O QCCD permite conservar baixas taxas de erro ao confinar poucos íons por zona (evitando cristais extremamente longos), ao custo de um overhead temporal para deslocamento dos íons.

Contudo, mesmo com QCCD, há um consenso de que escalar para muitas centenas ou milhares de qubits em um único dispositivo aprisionado será difícil. (Monroe et al., 2014) et al. argumentam que, além de ~1000 qubits, a complexidade de interconectar todas as zonas e orquestrar movimentos cresce enormemente, e a própria ótica de controle (focos laser individuais para cada íon) e eletrônica de grade de eletrodos se tornam gargalos de engenharia (Brown, Kim & Monroe, 2016). Assim, a segunda abordagem é adotar explicitamente uma arquitetura modular distribuída: ou seja, usar vários cristais de íons separados (módulos), cada qual contendo um número manejável de íons (por exemplo 50), e conectar esses módulos por interfaces fotônicas que distribuem entrelaçamento entre íons de módulos diferentes (Brown, Kim & Monroe, 2016). Nesse esquema, cada módulo age como um registrador quântico autônomo (às vezes chamado Elementary Logic Unit, ELU), capaz de executar portas locais internamente (via interações Coulomb locais) e também de entrelaçar um de seus qubits com um qubit de outro módulo através de fótons emitidos e interferidos em comum (Brown, Kim & Monroe, 2016). Tipicamente,

designa-se em cada módulo um ou alguns íons de comunicação especializados: esses íons podem ser excitados por pulsos laser para emitir fótons entrelaçados com seu estado quântico (por exemplo, a polarização ou frequência do fóton fica correlacionada com o estado interno do íon). Ao direcionar os fótons de dois módulos até um interferômetro com divisor de feixe 50/50 e detectá-los em coincidência, realiza-se uma medida de Bell nos fótons que projeta os dois íons emissores em um estado entrelaçado - evento este sinalizado pelos detectores (entrelaçamento heráldico). Esse protocolo de entrelaçamento remoto de nós quânticos por interferência de fótons já foi demonstrado com íons separados por metros a dezenas de quilômetros de fibra (Brown, Kim & Monroe, 2016), e com NV centers e outros sistemas.

A vantagem de íons aprisionados nesse contexto é sua natureza de qubits memória: coerências de segundos ou mais já foram observadas em transições de hiperfina de íons, permitindo que eles armazenem estados quânticos enquanto aguardam (até alguns segundos, se preciso) pelo sucesso de um protocolo de entrelaçamento remoto. Além disso, a interface de emissão de fótons por íons é relativamente bem controlada, especialmente se os íons estiverem acoplados a cavidades ópticas que direcionem a emissão. No entanto, as taxas de entrelaçamento remoto atualmente são limitadas por perdas de acoplamento e detecção. Em configurações típicas de pesquisa, coleta-se apenas poucos por cento da luz emitida por um íon e os detectores têm eficiência ~20%, resultando em probabilidades de sucesso muito baixas por tentativa ( $\sim 10^{-4}$ – $10^{-6}$ ). Consequentemente, mesmo com pulsos de excitação repetidos a MHz, as taxas efetivas de geração de pares entrelaçados entre módulos têm sido da ordem de dez a centenas de hertz. Por exemplo, (Monroe et al., 2014) et al. estimam uma taxa típica de ~100 Hz para a conexão entre dois módulos de íons com tecnologia atual (coleta de ~10% e detecção 20%). Em uma demonstração seminal, Hucul et al. (2015) usaram um interferômetro de fótons para entrelaçar íons em armadilhas separadas. A coleta eficiente de luz permitiu um entrelaçamento heráldico a  $4,5 \text{ s}^{-1}$ , com fidelidade de Bell cerca de 0,86.

Para superar essas limitações, esforços de engenharia se concentram em aumentar a eficiência das interfaces fóton-íon - por exemplo, integrando espelhos ou guias de onda nas armadilhas (como cavidades nanofotônicas ao redor dos íons) para coletar uma fração maior dos fótons emitidos. Resultados recentes sugerem que cavidades ultrapequenas em fibras ópticas nanoscópicas podem elevar a cooperatividade fóton-íon a ponto de gerar pares Bell entre módulos de átomos neutros ou íons a taxas de kHz a 1 MHz. Com tais avanços, espera-se que a latência de operações remotas em arquiteturas modulares de íons diminua drasticamente.

Empresas líderes em íons aprisionados já estão adotando arquiteturas modulares híbridas. A IonQ utiliza atualmente cadeias lineares de íons (por exemplo, 11 ou 23 íons em seus sistemas comerciais) com conectividade totalmente por modos vibracionais dentro da cadeia. Para escalar além disso, a IonQ planeja interligar múltiplas cadeias via fótons – em linha com a proposta de Monroe et al. (2014). Cada cadeia atuaria como um módulo, e íons de comunicação em diferentes módulos seriam entrelaçados por interferência de fótons, criando essentially uma única máquina maior distribuída espacialmente. Já a Honeywell/Quantinuum seguiu inicialmente a rota QCCD: seu protótipo System Model H1 implementou vários segmentos de armadilha em chip criogênico, entre os quais íons são continuamente transportados para efetuar um grafo completo de interações. Em 2021, o grupo da Honeywell reportou a integração completa da arquitetura QCCD, incluindo demonstração de uma porta CNOT teleportada (realizada entre íons sem interação direta, via medida intermediária) dentro de seu processador de 6 qubits (Pino, J. M. et al., 2021). Essa “porta teletransportada” é um exemplo de teleporte de porta lógica (ver Seção 4) e atingiu alta fidelidade, contribuindo para um resultado de volume quântico 64 (Pino, J. M. et al., 2021). A Quantinuum planeja combinar o QCCD com enlaces fotônicos entre múltiplos chips no futuro, unindo o melhor dos dois mundos – alta conectividade interna via shuttling e alcance global via entrelaçamento remoto.

Em suma, íons aprisionados são naturalmente adequados a arquiteturas modulares devido à sua estabilidade como memórias quânticas e à facilidade de gerar entrelaçamento fotônico heráldico entre módulos. O desafio reside em melhorar as



taxas de conexão e integrar de forma confiável muitos módulos sem perder a fidelidade. Krutyanskiy et al. (2023) demonstraram entrelaçamento heráldico de dois íons separados por 230 m, alcançando fidelidade aproximada de 88 % e probabilidade de sucesso de  $4 \times 10^{-5}$ . Esse trabalho não trata de radiação cósmica; para erros correlacionados causados por múons e raios  $\gamma$  em transmons, ver Li et al. (2025). Com avanços em óptica integrada e correção de erros, espera-se que redes de módulos de íons possam escalar para dezenas ou centenas de módulos, cada qual com dezenas de qubits – atingindo, no agregado, milhares de qubits logicamente utilizáveis.

### 3.1.3 Átomos Neutros

Átomos neutros em armadilhas ópticas vêm despontando como uma plataforma promissora tanto para computação quântica quanto para redes quânticas. Diferentemente dos íons, átomos neutros não carregados interagem menos uns com os outros naturalmente, mas podem ser dispostos em arranjos de grades ópticas ou pinças laser (optical tweezers) e induzidos a interagir fortemente via estados de Rydberg (átomos altamente excitados com grande momento dipolar) ou mediadores fotônicos em cavidades. Arranjos de átomos neutros já atingiram escalas de 100–300 átomos em demonstrações de simulação quântica e computação digital variacional (empresas como QuEra e Pasqal lideram esses esforços). Contudo, a maioria desses sistemas opera todos os átomos em um mesmo conjunto monolítico – uma única nuvem ou matriz. Assim como nos casos anteriores, surgem limitações práticas ao se tentar aumentar indefinidamente o número de átomos em um único arranjo: a estabilidade das armadilhas ópticas, o controle individual (é necessário um feixe de laser por átomo para endereçamento seletivo), bem como efeitos de decoerência coletiva (decoerência de um átomo pode espalhar fótons que perturbam vizinhos etc.), tendem a piorar com mais átomos.

Analogamente, a ideia de arquitetura modular aplica-se também aos átomos neutros. A proposta é utilizar múltiplos processadores de átomos neutros interligados por fotônica, em vez de um único conjunto extremamente grande (Sunami, S. et al., 2025). Cada módulo poderia ser, por exemplo, um registro de átomos neutros armadilhados em 2D (como as matrizes de tweezers) de tamanho moderado (talvez 50–200 átomos). Dentro de cada módulo, as interações quânticas podem ocorrer via excitações Rydberg de dois átomos vizinhos (realizando portas de dois qubits), ou via acoplamento de todos os átomos a um modo comum de cavidade (implementando portas coletivas). Já entre módulos, imagina-se conectar átomos específicos de cada módulo através de enlaces fotônicos semelhantes aos discutidos para íons – isto é, fazendo certos átomos emitirem fótons entrelaçados e interferindo-os para gerar estados emaranhados entre módulos. Átomos neutros possuem a vantagem de disporem de transições ópticas convenientes, inclusive em comprimentos de onda telecom (por exemplo, o itérbio-171 neutro -  $^{171}\text{Yb}$  - pode ser acoplado a interfaces fotônicas compatíveis com fibras), e tempos de vida de estado quântico longos (átomos em estados de storage como níveis hiperfinos ou estados de spin em cavidades podem reter coerência por segundos). Isso os posiciona como excelentes nós de memória e processamento em redes quânticas de longa distância (Covey, Weinfurter, Bernien, 2023).

Um elemento crítico para ligar módulos de átomos neutros é aumentar a probabilidade de sucesso do entrelaçamento remoto. Uma linha de pesquisa é integrar os átomos em cavidades ópticas de alto fator de Purcell, de modo que a emissão espontânea preferencialmente saia por modos bem definidos e acoplados a fibras. Sunami et al. (2025) analisaram uma arquitetura de múltiplos processadores de átomos neutros conectados por fibras acopladas a microcavidades de fibra (nanofibras). Eles mostram que cavidades com alta finesse e volume módico milimétrico podem acoplar  $>100$  átomos simultaneamente ao modo óptico, atingindo cooperatividade acima de 100 por átomo para transições no comprimento de onda de telecom do itérbio (Sunami, S. et al., 2025). Isso permitiria gerar pares de átomos entrelaçados entre módulos com taxa na ordem de megahertz, mantendo alto grau de indistinguibilidade e sincronismo. Com essa taxa, múltiplos módulos de átomos

neutros poderiam operar em conjunto praticamente sem sacrificar velocidade de computação, diferentemente do cenário atual de enlaces lentos (Sunami, S. et al., 2025). Ou seja, a rede fotônica não seria o gargalo – um avanço considerável.

Do ponto de vista experimental, já foram demonstrados alguns passos: entrelaçamento remoto de átomos únicos em cavidades separadas foi obtido, bem como teleporte quântico de estados entre nuvens de átomos. Covey et al. (2023) revisam os avanços que apontam para redes de átomos neutros como plataformas integradas de computação e comunicação. Eles destacam que arrays de átomos individuais controlados podem tanto executar processamento local de forma escalável quanto gerar entrelaçamento remoto eficiente, reunindo os pré-requisitos para redes quânticas complexas (Covey, Weinfurter, Bernien, 2023). A Figura 1 do trabalho deles ilustra uma possível rede de longa distância com nós de átomos neutros em cavidades e repetidores quânticos a cada ~50 km, onde qubits de comunicação (em vermelho) estabelecem vínculos entre nós e qubits de dados (verde) armazenam e processam informação localmente.

Em resumo, a modularidade para átomos neutros está no horizonte, com propostas concretas de multiprocessadores quânticos fonicamente interligados. Os desafios incluem desenvolver cavidades ópticas integradas aos arranjos de átomos, multiplexar muitos canais de comunicação paralelos para elevar throughput (vários pares de átomos diferentes gerando entrelaçamento simultaneamente entre dois módulos) e manter a coerência dos átomos durante as operações remotas. Como vantagem, átomos neutros operam em temperatura ambiente ou pouco acima (resfriados por laser, mas não requerem 10 mK como qubits supercondutores) e a infraestrutura de fibras telecom para conectá-los é madura. Assim, projeta-se que redes híbridas de átomos neutros possam escalar para dezenas de nós em escala de laboratório (distâncias de metros a salas distintas), e com repetidores, talvez em escala metropolitana, conjugando computação quântica distribuída com comunicações quânticas.

### 3.1.4 Qubits Fotônicos

Qubits fotônicos – em particular fótons individuais ou modos em estados quânticos de luz – formam uma plataforma naturalmente adequada a arquiteturas distribuídas, afinal fótons são as entidades portadoras de informação em redes. A computação quântica fotônica pode ser realizada de duas formas principais: com fótons discretos (onde cada fóton representa um qubit, codificado por polarização, caminho, etc.) ou com variáveis contínuas (estados quânticos de modos ópticos, como estados comprimidos). Ambas as abordagens se beneficiam do fato de fótons viajarem a grandes distâncias com perdas relativamente baixas em fibras ópticas e não requererem criogenia – sistemas fotônicos normalmente operam em temperatura ambiente, salvo os detetores que podem ser supercondutores para alta eficiência.

Em contrapartida, fótons não interagem diretamente uns com os outros de forma simples. Implementar portas lógicas entre qubits fotônicos tipicamente requer medidas projetivas (medição de alguns fótons com resultado heráldico) ou materiais não lineares. A célebre abordagem de KLM (Knill, Laflamme & Milburn) mostrou que portas universais fotônicas são possíveis usando apenas óptica linear e detecção, ao custo de preparações ancila e sucesso probabilístico. Hoje em dia, a estratégia predominante é construir estados de cluster fotônicos entrelaçados multimodo, que servem como recurso para computação quântica universal via medições adaptativas (computação quântica one-way). Nesse paradigma, a ênfase está em gerar um grafo grande de entrelaçamento entre muitos modos ópticos – o qual pode ser concebido como um “módulo” lógico – e então usar medição para efetuar as portas.

A modularidade na computação fotônica ocorre em múltiplas frentes. Uma é a modularidade temporal: em vez de usar muitos componentes distintos para fótons simultâneos, experimento recentes (como o processador Borealis da Xanadu) empregam multiplexação no tempo, guiando pulsos de luz através de linhas de atraso e interferômetros reconfiguráveis para construir um cluster state de grande tamanho com relativamente poucos recursos físicos (Madsen, Lars S. et al., 2022). No Borealis, por exemplo, ~3 km de fibra ótica são usados como memória de delay, permitindo que 216 modos temporais



interajam em um circuito de poucas portas squeezers e beam-splitters, resultando em um cluster de 216 modos ópticos (variáveis contínuas) entrelaçados (utilizado para amostrar distribuições difíceis de simular classicamente) (Madsen, Lars S. et al., 2022). Essa auto-conexão temporal dos modos pode ser vista como conectar diferentes “fatias de tempo” do processador – um tipo de rede quântica interna.

Outra frente é a modularidade espacial/macrosscópica: montar vários chips fotônicos ou racks de equipamentos e interconectá-los por fibra para escalar o número de modos/qubits processados. Como não há um limite claro de distância para conectar fotonicamente (além da atenuação da fibra, que a 1550 nm é  $\sim 0,2$  dB/km), pode-se imaginar agrupamentos modulares em escala de datacenter. De fato, a empresa Xanadu anunciou em 2025 seu sistema Aurora, descrito como “o primeiro computador quântico fotônico universal” que utiliza uma arquitetura modular de quatro racks interconectados por 13 km de fibra, operando a temperatura ambiente (Xanadu, 2025). Cada rack contém 12 qubits lógicos fotônicos formados por 35 chips fotônicos integrados, e a comunicação entre racks é inteiramente óptica. Esse design foi concebido para escalar linearmente: a Xanadu projeta que aumentando para milhares de racks semelhantes, pode-se atingir milhões de qubits fotônicos interconectados (Xanadu, 2025). O CEO da Xanadu argumenta que “a fotônica é a forma mais natural de tanto computar quanto interconectar”, ressaltando que a empresa já demonstrou a escalabilidade de networking necessária – restando focar agora em melhorar o desempenho quântico (reduzir perdas, implementar correção de erros).

A modularidade fotônica apresenta vantagens singulares: não há necessidade de sincronizar relógios quânticos locais (os fótons podem levar seu estado diretamente), não há ruído térmico relevante em frequências ópticas, e componentes padronizados de telecomunicações podem ser usados (fibras, comutadores ópticos, amplificadores paramétricos para geração de estados etc.). Por outro lado, o maior desafio é a perda óptica. Cada conexão, junção ou detecção introduz perda, e diferentemente de qubits materiais, não há como recuperar um estado fotônico perdido (não se pode clonar qubits para amplificar sinal). Assim, modularidade fotônica escalável requer altíssima eficiência em todos os estágios: fontes de fótons quase determinísticas, acoplamento de quase 100% entre chips e fibras, detetores com eficiências  $>99\%$ , além de esquemas de correção de erros adaptados a perdas (como códigos de detecção de perda ou repetidores quânticos para regenerar entanglement). Xanadu indica que parte de sua estratégia é reduzir a perda óptica por design, trabalhando em conjunto com foundries de fotônica em silício para otimizar cada chip e interconexão (Xanadu, 2025). De certo modo, a modularidade permite também contornar a perda: ao dividir o processamento em módulos, cada módulo pode produzir resultados intermediários que são medidos/ressupridos ao conectar com o próximo módulo, agindo como um repeater dentro do computador.

Concluindo, qubits fotônicos se beneficiam intrinsecamente de arquiteturas em rede. Enquanto plataformas materiais buscam como levar seus qubits de um módulo a outro, a fotônica já opera nativamente em uma rede de fibras e circuitos ópticos. A questão passa a ser como orquestrar um cálculo distribuído confiável com esses módulos. A resposta envolve medir e teleportar estados lógicos entre blocos cluster, usar feed-forward ultrarrápido (eletrônica de sincronização) e eventualmente integrar memórias quânticas materiais para armazenar fótons temporariamente (no caso de necessidade de correção de erros quânticos). Em última análise, se os desafios de perdas e não-linearidades forem vencidos, a fotônica modular oferece um caminho para datacenters quânticos em larga escala, com milhões de modos entrelaçados espalhados em diversos módulos interconectados (Xanadu, 2025).

### 3.1.5 Conectividade entre Módulos e Limitações Físicas

A conexão eficaz entre módulos quânticos é o cerne das arquiteturas modulares. Diferentes plataformas utilizam diferentes meios físicos de interconexão, mas em todos os casos o objetivo é distribuir entrelaçamento ou estados quânticos

coerentes entre módulos separados. Nesta seção, discutimos os principais métodos de conectividade e as limitações físicas associadas.

**Interfaces Fotônicas (Ópticas):** Conforme visto, para íons aprisionados, átomos neutros e muitos outros sistemas estacionários, a via preferencial de conexão é por meio de fótons. Tipicamente, cada módulo possui um ou mais qubits de comunicação que podem emitir fótons entrelaçados com seu estado (seja através de emissão espontânea num nível adequado, seja convertendo um estado de spin em fóton por meio de uma cavidade). Dois módulos trocam esses fótons – seja enviando-os a um módulo intermediário de interferência, seja emitindo um fóton de um módulo direto para um detector no outro módulo (no caso de interconexões diretas por fibra entre módulos). O entrelaçamento é estabelecido quando os fótons são correlacionados adequadamente por interferometria e detecção.

A principal limitação física aqui é a perda óptica. Fótons podem ser perdidos no acoplamento do emissor à fibra (ou meio livre), na propagação (atenuação na fibra, que a 1550 nm é cerca de 0,2 dB/km, significando ~95% de transmissão a cada km), e na detecção (detetores típicos têm eficiência de 50–95%). Se a probabilidade total de sucesso por tentativa de entrelaçamento é muito baixa (p. ex. 1% ou 0,1%), então muitas tentativas são necessárias em média para conseguir um par entrelaçado. Isso impõe uma latência nas operações inter-módulos, que pode ser significativa comparada às operações locais. Por exemplo, suponha uma taxa de tentativa de 1 MHz e sucesso de 1% por tentativa – em média tem-se 10 kHz de pares entrelaçados por segundo. Em 1% do tempo (10  $\mu$ s) se consegue, mas às vezes leva 1 ms (1000 tentativas) para conseguir um par. Em comparação, portas locais ocorrem em microssegundos ou menos, criando um desbalanço.

Para contornar isso, uma solução é usar múltiplos enlaces em paralelo e memória quântica para armazenar pares bem-sucedidos enquanto outros ainda estão tentando. Assim, pode-se manter um fluxo constante de entanglement. Ainda assim, a velocidade da luz entra como fator: se módulos estão distantes (digamos 1 km), só o tempo de ida e volta de um fóton é ~6,6  $\mu$ s, o que impõe latência mínima mesmo que o sucesso fosse imediato. Em escalas maiores (por exemplo, interconectando cidades), latências de milissegundos a dezenas de ms aparecem (pois 1000 km equivalem ~5 ms unidirecional em fibra, sem contar esperas). Logo, arquiteturas modulares intracampus (mesma sala ou prédio) podem ter latência óptica desprezível (~ns a  $\mu$ s), enquanto arquiteturas interurbanas exigirão considerar latências de comunicação evidentes.

Outra limitação fotônica é a estabilidade de fase e sincronização. Para interferir fótons de dois módulos de forma indistinguível, é preciso controlar com precisão relativa a diferença de caminho ótico entre eles dentro de uma fração do comprimento de onda (para íons com diferença de frequência de hiperfina ~12.6 GHz, basta estabilizar a diferença de caminho a ~1 mm de precisão (Brown, Kim & Monroe, 2016)). Isso requer estabilização ativa de fibras (compensação de expansão térmica, vibração) ou uso de esquemas de interferometria unidirecional insensíveis a fase (ex.: codificação de tempo). Em links longos, a deriva de fase com a temperatura do ambiente pode exigir feedback frequente. Nos experimentos, isso tem sido feito com sucesso em escalas de até 50 km usando fibras estabilizadas.

**Interfaces de Micro-ondas (SC):** No caso de qubits supercondutores, que operam em micro-ondas (~5–10 GHz), uma opção de interconexão é usar cavidades ressonantes compartilhadas ou linhas de transmissão. Por exemplo, dois chips supercondutores em fridges separados poderiam ser ligados por um cabo coaxial supracondutor, formando efetivamente um modo ressonante comum que mediatiza entrelaçamento entre qubits de cada lado. Já se demonstrou entrelaçamento de qubits supercondutores separados por vários metros através de um cabo coaxial com amplificadores de baixo ruído, usando protocolos de medida (Cantori, Simone; Pfaffhauser, Marcel; Bäumer, Elisa; Scafrimuto, Fabio; Davis, Robert, 2025). Entretanto, cabos têm perda (~0,1 dB/m mesmo a 20 mK e frequências GHz), e mais crítico, precisam sair do ambiente completamente isolado de um chip e entrar em outro – isso pode introduzir ruído térmico se não for bem isolado, além de limitar escalabilidade (cada conexão é um cabo dedicado). Até o momento, a maior parte das conexões SC bem sucedidas

envolveu módulos muito próximos (no mesmo cryopackage), permitindo couplers diretos com perdas mínimas (Niu, Jingjing et al, 2023).

Uma solução ativa em pesquisa para SC é a conversão quântica de frequência: converter o estado de um qubit supercondutor (fóton de micro-onda) para um fóton óptico, via meio não linear (como optomecânica ou cristais não lineares acoplados a ressonadores SC). Se essa conversão pudesse atingir eficiências altas (>50% idealmente), então qubits SC herdariam as vantagens da distribuição por fibra óptica (longa distância). Contudo, alcançar essa conversão sem introduzir ruído excessivo é um grande desafio – eficiências atuais estão na casa de poucos por cento.

**Acoplamento Direto de Matérias:** Em princípio, módulos que estejam muito próximos (micrômetros a centímetros) podem interagir por campos diretos: por exemplo, dois módulos de íons aprisionados colocados em armadilhas vizinhas podem acoplar via campo elétrico mútuo se suficientemente próximos, ou qubits de estado sólido (como NV centers) em chips distintos podem interagir via acoplamento dipolar se a separação for muito pequena. No entanto, essas abordagens não escalam para distâncias mais longas e são altamente específicas (exigem proximidade extrema), de modo que a interconexão fotônica ou de micro-ondas propagante tende a ser preferida.

Resumindo, limitações físicas chaves nas interconexões modulares incluem: - **Perda e Eficiência:** determina a taxa de sucesso e, portanto, a latência das operações remotas. - **Velocidade de Propagação:** impõe latência mínima e influencia a arquitetura de controle (ex.: tempos de espera para feed-forward). - **Decoerência durante espera:** qubits nos módulos precisam reter coerência durante o tempo de estabelecimento do entrelaçamento – um ponto forte dos íons/neutros, mas ponto fraco para qubits SC (que tipicamente decoerem em microssegundos ou milissegundos). Assim, se qubits SC tiverem de esperar 1 ms por um entrelaçamento remoto, muitos podem perder coerência; isso sugere usar entrelaçamento remoto principalmente para qubits robustos (memórias) ou aplicar correção de erros mesmo durante a espera. - **Crosstalk e Ruído entre Módulos:** quando interligamos módulos, abrimos canais por onde ruído também pode se propagar. Por exemplo, luz espalhada de um módulo pode chegar em outro se fibras não forem bem isoladas, ou flutuações de corrente em um cabo coaxial podem introduzir ruído em ambos módulos SC. O desenho cuidadoso de filtros, isoladores e blindagens é necessário. - **Complexidade de Comutação:** se houver muitos módulos e quisermos conectá-los de forma flexível (não apenas ponto-a-ponto fixo), precisamos de comutadores quânticos. Na fotônica, isso significa consertos ópticos reconfiguráveis (espelhos ou switches MEMS) que direcionem fótons de um módulo para outro conforme necessário. Já se demonstrou a ideia de optical cross-connect para íons (Brown, Kim & Monroe, 2016), mas inserir tal comutador adiciona perda extra. Existe um balanço entre topologia da rede (totalmente conectada vs. comutação dinâmica vs. topologia fixa tipo anel, malha, etc.) e a simplicidade/eficiência. - **Interoperabilidade de Tecnologias:** em arquiteturas heterogêneas (e.g., conectar um módulo de íons a um de supercondutor via fotônica), adiciona-se mais camadas (como conversores de frequência, interfaciamento de protocolos) que podem degradar fidelidade. Até agora, a maioria das implementações foca em módulos idênticos (homogêneos) interconectados.

Ao projetar arquiteturas modulares, é essencial quantificar essas limitações e incorporá-las nos modelos de computação e correção de erros. Por exemplo, a tolerância a falhas global dependerá não só dos erros de porta locais, mas também da taxa/erro de entrelaçamento remoto. Um enlace com 10% de erro (90% fidelidade) pode ser aceitável se usado raramente e purificado, mas se usado intensivamente exigirá overhead grande de correção. Assim, muitas propostas definem um “limiar de ruído de rede” – a fração máxima de erro no enlace entre módulos que ainda permite computação tolerante a falhas. Nickerson et al. calcularam um threshold de até ~13,3% de erro no enlace para um protocolo específico de correção (código de superfície)(Nickerson, Fitzsimons & Benjamin, 2014), mostrando que mesmo enlaces muito ruidosos podem servir, desde que abaixo desse limite e com purificação para melhorar fidelidade.

Em suma, conectar módulos quânticos envolve equilibrar físico (perda, latência) com lógico (protocolos para

contornar probabilidade de falha). Tecnologias emergentes como repetidores quânticos (ver próxima seção) e buffers quânticos podem mitigar algumas limitações, mas sempre à custa de maior complexidade.

### 3.1.6 Redes Quânticas: Entrelaçamento Distribuído, Repetidores e Teleporte de Portas Lógicas

Quando módulos quânticos são interligados por canais de comunicação, formamos o que se pode chamar de uma rede quântica. Redes quânticas permitem distribuir estados entrelaçados entre diferentes nós (processadores) e, assim, habilitar novas capacidades: desde comunicação quântica segura até computação quântica distribuída e sensoriamento coordenado (Covey, Weinfurter, Bernien, 2023). Nesta seção, exploramos três conceitos-chave nesse contexto: entrelaçamento distribuído entre módulos, o uso de repetidores quânticos para ampliar o alcance de redes, e o teleporte de operações lógicas (portas quânticas teleportadas) como mecanismo para realizar computação entre módulos distantes.

**Entrelaçamento Distribuído:** A base de qualquer rede quântica funcional é a capacidade de gerar entrelaçamento (correlações quânticas não locais) entre qubits localizados em nós diferentes. Conforme discutido, isso geralmente é alcançado por protocolos heráldicos envolvendo um recurso móvel (fóton) que viaja de um nó a outro ou a um ponto intermediário. Quando dois nós compartilham um par de qubits entrelaçados (um em cada nó), eles passam a ter um recurso poderoso: podem correlacionar instantaneamente medições, podem implementar teleportes de estado e, fundamentalmente, podem atuar como se esses qubits fossem adjacentes em um circuito. Por exemplo, suponha que o Módulo A e o Módulo B possuem um par de qubits  $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$  entrelaçados entre si (A com B). Esse estado pode ser consumido para, via protocolo de teleporte, transferir um estado arbitrário de um qubit adicional em A para algum qubit em B, comunicando apenas bits clássicos. Alternativamente, pode servir de “canal quântico” para executar uma porta lógica remota, como veremos adiante.

**Repetidores Quânticos:** Em redes quânticas de longa distância, a taxa de sucesso de entrelaçamento decai exponencialmente com a distância devido a perdas. Para superar distâncias maiores que a distância de atenuação (~ dezenas de km em fibra), foram propostos os repetidores quânticos: nós intermediários que dividem o link em segmentos menores, estabelecem entrelaçamento em cada segmento, e então aplicam swap de entrelaçamento para conectar segmentos (Covey, Weinfurter, Bernien, 2023). O swap de entrelaçamento é essencialmente a medida de Bell em qubits intermediários de segmentos adjacentes, que projeta os qubits das extremidades em um estado entrelaçado, estendendo assim o alcance. Por exemplo, se quisermos entrelaçar Alice e Bob a 1000 km, podemos colocar repetidores a cada 50 km; inicialmente criamos pares entrelaçados (Alice-Rep1, Rep1-Rep2, ..., RepN-Bob) em cada trecho de 50 km; então pedimos que cada repetidor faça uma medida conjunta em seus qubits de dois segmentos (por exemplo, Rep1 mede o qubit ligado a Alice com o qubit ligado a Rep2). Isso conectará Alice a Rep2 em um par entrelaçado (com probabilidade e redução de fidelidade dependendo de erros). Repetindo ao longo da cadeia, finalmente obtém-se Alice e Bob entrelaçados, sem nunca ter enviado um fóton diretamente entre eles. Repetidores requerem armazenar qubits temporariamente (memórias quânticas) enquanto esperam todas as etapas terem sucesso, e idealmente incluem purificação de entrelaçamento em etapas para elevar a fidelidade final metros (Sunami, S. et al., 2025). Portanto, repetidores quânticos são mini-computadores quânticos por si próprios, com qubits de memória, qubits de comunicação e capacidade de executar portas de dois qubits e medições locais.

Embora repetidores sejam mais relevantes para comunicações de longa distância, também podem ser úteis dentro de um computador quântico modular muito grande. Imagine uma arquitetura modular com centenas de módulos: em vez de tentar entrelaçar diretamente um módulo A com Z muito distante, poderíamos entrelaçar A com B, B com C, ..., Y com Z localmente (cada vizinho próximo), e então usar swaps para conectar A–Z. Isso é análogo a redes clássicas roteadas. Porém, no contexto de um único computador quântico, possivelmente podemos permitir um cross-connect óptico que interliga quaisquer dois módulos diretamente a curta distância com perdas moderadas (Brown, Kim & Monroe, 2016), evitando precisar de muitos

saltos. Ainda assim, se o número de módulos for enorme, camadas hierárquicas de rede (e repetição) podem surgir como solução arquitetônica.

**Teleporte de Portas Lógicas:** Talvez o aspecto mais sutil e distintivo de computação distribuída seja a habilidade de realizar operações quânticas entre qubits que não estão no mesmo local físico. Suponha que temos um qubit no módulo A e outro no módulo B, e desejamos aplicar uma porta lógica de dois qubits (por exemplo, CNOT ou CZ) entre eles. Se A e B compartilham um par entrelaçado pré-estabelecido, existe um protocolo que realiza essa porta usando apenas operações locais e comunicação clássica – essencialmente uma extensão do teleporte de estado, conhecida como teleporte de porta ou porta teleportada.

A ideia, descrita em trabalhos de Gottesman e Chuang no fim dos anos 1990, é que certas portas quânticas (particularmente as da classe Clifford e algumas não-Clifford adaptadas) podem ser efetuadas combinando o estado entrelaçado e medidas apropriadas. Por exemplo, para aplicar uma porta CNOT remota, podemos seguir um protocolo: (1) entrelaçar um qubit auxiliar do módulo A com o qubit alvo de A; (2) entrelaçar um qubit auxiliar do módulo B com o qubit alvo de B; (3) usar o par entrelaçado entre módulos como canal; (4) fazer medições conjuntas (Bell) envolvendo os qubits auxiliares e comunicar os resultados para aplicar correções condicionais. Em essência, o custo de uma porta remota é consumir um par de entrelaçamento pré-compartilhado e realizar duas medições de Bell e algumas correções de 1-qubit via comunicação clássica. O resultado final é que a mesma transformação unitária ocorreria entre os qubits alvo como se eles interagissem diretamente.

Um exemplo concreto está no experimento da Honeywell (Quantinuum) citado: eles realizaram uma CNOT teleportada entre dois íons que não interagiam diretamente na mesma armadilha, usando medições de meio de circuito e comunicação clássica (Pino, J. M. et al., 2021). A equipe da Honeywell demonstrou uma porta CNOT teleportada na arquitetura QCCD e obteve fidelidade média de processo  $\approx 0,90$  ( $F_{\text{med}} \geq 0,899(6)$ ), com volume quântico  $2^6 = 64$ . De maneira geral, teleporte de portas lógicas requer entrelaçamento de alta fidelidade disponível e baixa latência de comunicação clássica (para aplicar as correções rapidamente sem deixar qubits decoerirem). Em escalas de laboratório, a latência clássica (poucas centenas de ns via cabos) é desprezível frente ao tempo de porta quântica (usualmente  $\mu\text{s}$  ou mais), então é viável.

Importante notar que, uma vez que possamos teletransportar portas arbitrárias, a separação física torna-se irrelevante para o poder computacional – vários módulos conectados por pares EPR adequados e comunicações suficientemente rápidas constituem um único computador quântico distribuído. Entretanto, cada porta teleportada consome recursos (pares entrelaçados) e introduz pontos potenciais de falha (se o par EPR não tiver fidelidade perfeita, a porta sofrerá erro). Assim, arquiteturas modulares otimizam para minimizar o número de portas remotas necessárias, localizando o máximo de circuito possível dentro de módulos locais e usando portas teleportadas apenas quando inevitável (por exemplo, para acoplar subcircuítos independentes ou interagir qubits que residem em módulos diferentes por restrição de tamanho).

Outro uso fundamental do teleporte de operações é na implementação de correção de erros quânticos distribuída. Em algumas propostas, cada módulo pode manter um qubit lógico codificado (por exemplo, um patch de código de superfície) e operações entre qubits lógicos em módulos distintos são feitas por teleporte lógico – consumindo estados entrelaçados de alta dimensão chamados magic states ou similares. Nesses casos, a rede quântica entre módulos suporta a manutenção de um registrador lógico distribuído com tolerância a falhas.

Resumindo esta seção: redes quânticas fornecem a infraestrutura que conecta módulos, e através delas podemos: - Compartilhar entrelaçamento (pré-requisito para qualquer operação não-local). - Estender alcance com repetidores quânticos (superando perdas exponenciais). - Executar computação distribuída via teleporte de estados e portas, permitindo que múltiplos módulos colaborem em um algoritmo unificado.

No contexto de arquiteturas modulares quânticas, muitas vezes falamos de uma rede quântica de curta distância (intrametropolitana ou dentro de um laboratório) com a finalidade de montar um computador quântico paralelo. Essa rede pode ser vista como um interconexômetro quântico do computador, análogo a barramentos em arquiteturas clássicas multiprocessadas.

### 3.2 Modelos de Custo e Latência para Operações Distribuídas

Para avaliar o desempenho de uma arquitetura modular, é crucial modelar quantitativamente o custo (em recursos físicos) e a latência (tempo de execução) das operações que envolvem múltiplos módulos. Enquanto em um processador monolítico todas as portas têm custo uniforme (em tempo e erro) até certa aproximação, em um computador quântico distribuído as operações não são todas equivalentes: portas locais em um módulo típico são rápidas e de alta fidelidade, ao passo que portas remotas (que envolvem comunicação quântica entre módulos) tendem a ser mais lentas e propensas a erros. Assim, arquitetos de sistemas quânticos devem levar isso em conta ao distribuir algoritmos entre módulos e ao determinar quantos módulos são necessários para uma dada tarefa.

Um parâmetro fundamental é a razão de velocidades ( $R = T_{\text{remota}} / T_{\text{local}}$ ), em que  $T_{\text{remota}}$  é o tempo médio para executar uma operação de dois qubits entre módulos e  $T_{\text{local}}$  é o tempo de uma operação local de dois qubits.

No caso de íons aprisionados,  $T_{\text{local}} \approx 10 \mu\text{s}$  (porta mediana entre dois íons no mesmo cristal), enquanto  $T_{\text{remota}}$  depende da taxa de entrelaçamento fotônico. Se a taxa de conexão for 100 Hz (como na estimativa de (Monroe et al., 2014) et al. para tecnologia atual), então  $T_{\text{remota}} \approx 10 \text{ ms}$  por tentativa bem-sucedida. Mesmo supondo melhora para 1 kHz ( $\approx 1 \text{ ms}$ ), obtém-se  $R \approx 10^2\text{--}10^3$ . Isso significa que, nesse exemplo, portas remotas são de 100 a 1000 vezes mais lentas que portas locais.

Para qubits supercondutores multi-chip de Gold et al. (2021),  $T_{\text{local}} \approx 50 \text{ ns}$  e  $T_{\text{remota}}$  era da mesma ordem ( $\approx 100 \text{ ns}$ , graças ao acoplamento flip-chip rápido) (Gold. A., et al, 2021). Assim,  $R \approx 2$  nesse sistema — praticamente transparente. Assim, diferentes plataformas apresentarão diferentes penhascos de latência entre local e remoto.

Modelos de custo computacional para algoritmos distribuídos incorporam essa diferença ponderando o custo das portas remotas. Por exemplo, ao compilar um circuito quântico para uma arquitetura modular, pode-se inserir “esperas” ou “no-ops” correspondentes à latência de comunicação. Em alguns casos, se a latência for muito grande, pode ser melhor replicar certos subcálculos em múltiplos módulos para evitar trocas intensas de qubits (similar a enviar código ao dado em vez de dado ao código).

Outro aspecto do custo é o consumo de recursos físicos para entrelaçamento remoto. Cada tentativa de entrelaçamento fotônico entre módulos envolve desperdiçar (quando falha) estados quânticos preparados, fotodetetores acionando etc. Protocolos de purificação de entrelaçamento aumentam a fidelidade ao custo de usar múltiplos pares brutos para obter um par refinado. Nickerson et al. mostraram que sob 98% de perda de ftons, ainda seria possível obter estabilizadores de código de superfície de alta fidelidade a  $\sim 1 \text{ kHz}$  consumindo entanglement purification intensivamente (Nickerson, Fitzsimons & Benjamin, 2014). O custo aí é que para cada par utilizável, gastam-se muitas tentativas de pares ruins. Assim, define-se uma taxa de entropia por enlace: quantos bits de entropia (erros) precisam ser expurgados via purificação e qual overhead isso implica em número de tentativas.

Em termos de latência, podemos distinguir: - Latência de geração de EPR: tempo para estabelecer um par entrelaçado entre módulos (pode incluir várias tentativas). - Latência de porta remota: inclui latência de EPR + tempo de medição e comunicação clássica de correções. - Latência de repeater (se houver): inclui armazenar qubits por tempo adicional até que segmentos vizinhos estejam prontos, etc.

Se a arquitetura implementa operações paralelas (por exemplo, tenta entrelaçar em vários canais simultaneamente),



parte do tempo morto pode ser mitigado. Em arquiteturas de íons e neutrões, geralmente há possibilidade de multiplexação espacial: vários pares de íons podem tentar entrelaçamento em paralelo com feixes distintos (Brown, Kim & Monroe, 2016). Nesse caso, a latência efetiva para obter  $N$  pares pode ser muito menor que  $N$  vezes a latência para 1 par – se houver hardware para paralelo.

O custo de comunicação clássica também deve ser mencionado. Ainda que a comunicação clássica viaje à velocidade da luz (ou sinal elétrico no cabo), em escalas de metros a diferença é ínfima, mas em escalas geográficas torna-se notável (milissegundos). Computação distribuída de larga área seria prejudicada por latências clássicas elevadas – assim como clusters clássicos distantes sofrem latência de rede. Porém, dentro de um computador modular num mesmo laboratório, o delay de  $\sim 3$  ns/m do sinal elétrico é comparável ou menor que tempos de porta de muitos qubits, logo é negligenciável. Por exemplo, a Honeywell conseguiu feed-forward de medida para correção de porta teleportada dentro do ciclo de clock sem problema (Pino, J. M. et al., 2021).

Outra métrica é o throughput ou taxa de operações por segundo que a arquitetura pode sustentar. Módulos múltiplos podem operar em paralelo independentemente, então idealmente o throughput escala com número de módulos para computação embarrassingly parallel. Porém, para problemas que requerem constante interação entre todos módulos (por ex., simular um sistema totalmente conectado), o throughput ficará limitado pelos enlaces mais lentos.

No design de algoritmos, um ponto central é minimizar a necessidade de comunicação quântica entre módulos, ou estruturar essa comunicação de forma hierárquica. Por exemplo, arquiteturas modulares às vezes se combinam melhor com protocolos de computação quântica distribuída específicos, como algoritmo de Shor distribuído, QFT distribuída, etc., onde parte do cálculo é local a cada módulo e apenas certos estágios envolvem agregação de resultados (similar a MapReduce clássico, mas quântico). Nesse contexto, quantificar quantas rodadas de comunicação quântica e clássica o algoritmo necessita fornece uma noção de escalabilidade. Se um algoritmo requer comunicação a cada passo, torna-se ineficiente em modular; mas se pode ser dividido em fases com pouca comunicação, a modularidade é vantajosa.

Finalmente, do ponto de vista de tolerância a falhas, os modelos de custo devem incorporar as taxas de erro. Uma operação remota malsucedida (por perda do par EPR ou falha na medição) pode ser detectada (heráldica indica quando falha, daí repetimos) ou pode levar a erros lógicos se não detectada. Isso entra no budget de erro do código de correção. Por exemplo, se cada porta remota falha 1% das vezes sem aviso, o código de correção deve suportar isso. Em contrapartida, se falha for heráldica (sabemos quando não deu certo), podemos simplesmente reintentar – o custo é tempo, não aumento da taxa de erro lógico (exceto pelo tempo extra expondo qubits à decoerência, que também conta).

Em suma, modelos analíticos e simuladores para arquiteturas modulares levam em conta:

- Probabilidade de sucesso por tentativa de enlace  $p$ .
- Tempo por tentativa  $t_{\text{trial}}$ , incluindo ida e volta de luz).
- Número de tentativas esperadas  $1/p$ .
- Fidelidade do par entrelaçado resultante.
- Tempo de processamento interno (incluindo possivelmente correção de erro local).
- Arranjos paralelos (quantos canais em paralelo).
- Latência clássica de sinalização.
- Recursos extras (qubits ancilla para purificação, etc).

Esses modelos orientam trade-offs: por exemplo, vale a pena purificar? Purificação usa duas ou mais conexões para aumentar fidelidade e, se for em paralelo, adiciona possivelmente menos latência do que esperar repetidamente por um par

perfeito. Se a purificação duplica a fidelidade mas quadruplica tempo, é preferível ou não? Depende se o protocolo de correção de erro requer aquela fidelidade para estar abaixo do threshold (Nickerson, Fitzsimons & Benjamin, 2014).

Trabalhos como o de Nickerson et al. e (Monroe et al., 2014) et al. vêm sugerindo que mesmo com enlaces muito lentos, consegue-se superar um computador clássico se o problema for adequado (por exemplo, fatoração de RSA não exige operações muito densamente conectadas, então um clock quântico de kHz poderia já superar HPCs em alguns casos). Em paralelo, se a tarefa for muito interativa entre módulos (como simular magneto totalmente conectado), latências altas podem estragar qualquer vantagem.

Assim, projetistas de arquiteturas modulares definem uma regra de bolso: mantenha a maioria das portas dentro de módulos (rápidas) e use portas remotas parcimoniosamente. E quando usar, assegure-se de ter entanglement sendo gerado continuamente em background para uso quando preciso, de modo a esconder latência (conceito de quantum buffering).

### 3.3 Escalabilidade e Desafios de Implementação Física

Embora arquiteturas modulares e redes quânticas prometam escalabilidade além do possível em sistemas monolíticos, sua realização prática enfrenta diversos desafios. Aqui destacamos alguns dos principais obstáculos técnicos e considerações de escalabilidade nas implementações físicas dessas arquiteturas.

**Integração e Engenharia de Sistemas:** Construir um computador quântico modular envolve integrar múltiplos sistemas complexos (vácuo, óptica, eletrônica de RF, fotônica, criogenia) de forma coesa. Cada módulo precisa ter desempenho quântico de ponta, mas também deve cooperar com outros módulos sem interferências deletérias. Por exemplo, no caso de íons aprisionados, módulos diferentes podem requerer lasers independentes; quando ligados em rede, os lasers de comunicação de cada módulo podem causar ruído nos vizinhos se não isolados. A engenharia do sistema deve garantir isolamento vibracional, eletromagnético e ótico apropriado entre módulos, ao mesmo tempo em que fornece canais de comunicação bem controlados entre eles.

**Sincronização Temporal:** Operações entre módulos requerem sincronização precisa de relógios e gatilhos. No nível físico, por exemplo, dois fótons de módulos distintos devem chegar simultaneamente a um detector para interferirem com sucesso. Isso exige calibrar trajetos com precisão de sub-nanosegundos. Em sistemas distantes, pode ser necessário usar lasers pulsados sincronizados ou relógios atômicos coordenados (desafios similares aos de distribuição de frequências em metrologia). As arquiteturas devem incluir mecanismos de calibração ativa contínua para manter a sincronização – por exemplo, enviando pulsos de teste e ajustando fases.

**Escalabilidade de Controle Clássico:** À medida que o número de módulos cresce, cresce também o volume de equipamentos clássicos de controle (geradores de pulso, fontes de micro-ondas, placas FPGA de sequenciamento, etc.). Uma máquina com 100 módulos de íons, cada módulo com 50 qubits, pode ter milhares de canais de laser/RF para controlar. Projetos atuais já demandam eletrônica de alta densidade e sofisticado software de orquestração (camadas de middleware quântico). Tornar isso escalável requer padronização de módulos (hardware e software) e arquitetura hierárquica de controle. A ABNT aponta preferencialmente bibliografias – no contexto de controle, livros de sistemas de controle quântico e artigos de implementação de grande escala (ex: sobre QSCOUT ou Q20:20 etc.) seriam referenciáveis, mas manteremos qualitativo.

**Fabricação Repetível e Rendimento:** Para alcançar dezenas ou centenas de módulos, é desejável que eles sejam idênticos ou pelo menos compatíveis. Isso favorece a adoção de plataformas chip-scale quando possível (e.g., armadilhas de íons micro fabricadas para cada módulo, chips fotônicos replicáveis). A técnica de IonQ e Honeywell de usar uma mesma espécie atômica em todos os módulos facilita homogeneidade. No caso supercondutor, montar muitos chips requer que cada um atinja certo patamar de fidelidade – talvez se fabrique 100 chips e escolha 50 melhores. Isso torna importante o yield de

fabricação. Arquiteturas modulares, por outro lado, toleram substituição de módulos defeituosos: se um módulo não funcionar, pode-se trocá-lo sem descartar todo o sistema (diferente de um mega-chip monolítico onde um defeito arruína tudo). Assim, do ponto de vista prático, modularidade melhora a escalabilidade ao relaxar a exigência de perfeição absoluta em cada componente (Niu, Jingjing et al, 2023).

**Comunicação com Baixa Taxa de Erro:** Em links quânticos, perdas muitas vezes podem ser detectadas (ausência de clique do detector), mas outros erros – como descoincidência de fótons ou decoerência de fase no meio – podem degradar a fidelidade sem sinalização óbvia. É crucial minimizar fontes de erro no enlace: vibrações que causem desestabilização de fase óptica, flutuação de comprimento de caminho óptico (por temperatura ou pressão), ruído nos detectores (dark counts) que podem falsear heralds, etc. Em protocolos com repetidores, garantir que memórias quânticas armazenem o estado sem decoerência durante o tempo de espera é outro desafio (ex.: memórias de estado de spin em NV centers para segurar entanglement 1s enquanto se entelha link seguinte).

**Correção de Erros Distribuída:** A correção de erros quânticos em larga escala será necessária para qualquer computador quântico útil, modular ou não. Contudo, nos sistemas modulares, surge a questão de como organizar códigos de erro entre módulos. Uma possibilidade é cada módulo hospedar certas partes de um código (por exemplo, um plano do código de superfície) e enlaces entre módulos atuarem como limites onde ocorre uma fusão de patches ou teleportação de síndromes. Isso é área de pesquisa ativa – implementar rotinas de QEC que tolerem também erros nos enlaces. Desafios incluem latência de comunicação que pode impedir completar ciclos de correção suficientemente rápido, e sincronização de clocks de correção entre módulos. Em geral, arquiteturas modulares ampliam a necessidade de engenharia de protocolos de QEC distribuídos robustos a perdas e atrasos.

**Casamento de Tecnologias Diferentes:** Alguns projetos vislumbram arquiteturas heterogêneas, onde módulos de tipos distintos colaboram. Por exemplo, poderia-se ter módulos de qubits supercondutores rápidos para processamento intensivo e módulos de íons aprisionados (ou NV centers) como nós de memória de longa duração – combinando o melhor de cada tecnologia. Esses módulos híbridos seriam conectados via interface fotônica ou até transdutores quânticos (micro-onda<->ótico). Embora conceitualmente atraente, isso impõe duplo desafio: dominar ambas tecnologias e a interface entre elas. Atualmente, cada tecnologia tem sua comunidade e complexidade; integrá-las sem comprometer desempenho é extremamente difícil. Por enquanto, a maioria das arquiteturas modulares propostas ficam dentro de uma mesma família de qubits e apenas usam fótons como “ponte universal” (fóton interage com ambos tipos). Um exemplo notável são redes de memória NV center (sólido) conectadas a qubits supercondutores via canais ópticos e transdutores; alguns experimentos isolados mostraram transferência de estado entre NV e transmon com sucesso moderado.

**Expansibilidade e Manutenção:** Em escala industrial, se imaginarmos um datacenter quântico com racks modulares (como o da Xanadu), uma consideração prática é como adicionar módulos incrementalmente e como realizar manutenção sem interromper todo sistema. Arquiteturas modulares devem, idealmente, suportar hot-swapping de módulos – retirar um módulo defeituoso e inserir outro sem desalinhar toda a óptica, por exemplo. Isso requer padronização mecânica e calibração rápida. Também, se o sistema tiver milhares de componentes (lasers, amplificadores, criocoolers), a confiabilidade individual importa: uptime do sistema dependerá de tempo médio entre falhas de componentes e da habilidade de consertar sem downtime extenso.

**Consumo de Energia e Infraestrutura:** Outro aspecto às vezes esquecido: muitos módulos significam possivelmente muitos lasers (para íons/átomos), muitos refrigeradores (para supercondutores), muitos detectores (em fotônica). O consumo energético e complexidade de suporte pode crescer linearmente ou pior. É preciso pensar em escalabilidade prática – por exemplo, supercondutores modulare: seria viável ter 100 refrigeradores a 10 mK funcionando em paralelo? Provavelmente não; preferível ter um refrigerador maior contendo vários módulos dentro. Isso leva a tentativas de ampliação do volume útil

de cryostats e multiplexação de linhas de controle.

**Crescimento do Grau de Conectividade:** Em arquiteturas muito grandes, pode não ser viável conectar todos com todos. Pode-se então adotar topologias de rede restritas (ex.: cada módulo conecta-se apenas aos  $k$  vizinhos mais próximos, formando uma rede de grau fixo). Isso limita algoritmos possíveis ou introduz overhead de roteamento (teleportar através de módulos intermediários). O design ideal de rede depende de suposições de uso. Por exemplo, redes lineares ou anéis de módulos minimizam hardware, mas aumentam distância de comunicação entre módulos distantes. Redes totalmente conectadas via um comutador central (estilo estrela) minimizam saltos de qubits, mas exigem um switch óptico capaz de direcionar qualquer par – isso pode se tornar um gargalo. Assim, escalabilidade requer pensar no “layout da rede quântica interna”: hierárquica (clusters de clusters), malha 2D, hipercubo, etc., cada qual com implicações.

Em suma, embora nenhuma barreira fundamental impeça arquiteturas modulares – de fato, elas são possivelmente o único caminho para milhões de qubits – há um caminho árduo de desafios de engenharia a superar. Até o momento, demonstrações experimentais têm se focado em poucos módulos (2 a 4 nós) e provado princípios de entrelaçamento remoto e portas teleportadas. Escalar para dezenas ou centenas de módulos exigirá avanços paralelos em diversas frentes: fontes de luz integradas, eletrônica de controle de alta densidade, detectores eficientes, packaging precisos, protocolos de calibragem autônoma e assim por diante. A comunidade de pesquisa já reconhece esses desafios e iniciativas de co-design de hardware/software (como a do projeto de Monroe et al. 2016) estão em andamento para abordá-los.

### 3.4 Estudos de Caso: Arquiteturas Modulares Reais e Propostas

Nesta seção, examinamos algumas arquiteturas modulares já implementadas em protótipos ou propostas por empresas líderes, ilustrando concretamente as abordagens discutidas. Focamos em três casos: a arquitetura modular de íons aprisionados da IonQ, a arquitetura QCCD e modular da Honeywell/Quantinuum, e a arquitetura fotônica modular da Xanadu. Essas empresas, cada uma em sua plataforma, fornecem visões complementares de como escalar computadores quânticos via modularidade.

#### 3.4.1 IonQ – Módulos de Íons Aprisionados interconectados por Fotônica

A IonQ, fundada por pesquisadores pioneiros em íons aprisionados (C. Monroe e J. Kim, entre outros), baseia sua tecnologia em armadilhas de íons lineares com lasers modulados individualmente para controlar qubits. As máquinas IonQ atuais (por exemplo, de 11, 23 qubits) utilizam cadeias únicas de íons  $\text{Yb}^+$  com conectividade total interna – qualquer par de qubits pode ser acoplado por portas MS (Mølmer-Sørensen) mediadas por fonons. Entretanto, para ir além de algumas dezenas de íons, a IonQ adota explicitamente a visão modular descrita neste artigo (Monroe et al., 2014). Em particular, Monroe e colegas delinearam uma arquitetura modular desde 2014, muitas vezes apelidada de “Quantum Lego” por sua natureza de blocos combináveis (Monroe et al., 2014).

No modelo IonQ, vários módulos de armadilha de íons seriam conectados por fibras ópticas e beam splitters. Cada módulo conteria talvez 50–100 íons manipuláveis localmente (número factível sem grande perda de fidelidade). Entre módulos, alguns íons especiais (denominados íons de comunicação) seriam usados para estabelecer entrelaçamento remoto via emissão de fótons e interferência. Uma patente e artigos da equipe de Monroe descrevem a utilização de unidades lógicas elementares (ELUs) de íons conectadas em rede fotônica não-bloqueadora, permitindo portas entre qubits arbitrários de módulos distintos (Brown, Kim & Monroe, 2016). Em termos práticos, a IonQ investe no desenvolvimento de fibras acopladas diretamente às armadilhas ou de sistemas de espelhos que colem fótons emitidos pelos íons. Cada módulo IonQ teria, por exemplo, um laser de pulso ultrarrápido para excitar o íon de comunicação a emitir um fóton; esse fóton entraria numa rede de

fibras e interferômetro que pode conectá-lo com um fóton de outro módulo, gerando entrelaçamento heráldico.

Um marco experimental relevante foi a demonstração, em 2015, do entrelaçamento entre dois íons em armadilhas separadas (no caso, separadas por 1 metro de fibra) com taxa de  $\sim 10$  Hz (Monroe, Christopher, 2014). Essa foi a primeira instância de arquitetura modular de íons operando: eles exibiram entanglement tanto intra-módulo (entre íons no mesmo chip via fonons) quanto inter-módulo (entre íons de chips diferentes via fótons) simultaneamente. A taxa de 10 Hz, embora baixa, foi limitada por ineficiências de acoplamento e detecção; a fidelidade alcançada ( $\sim$  entropia próxima de um estado puro) provou viabilidade, e os autores apontaram caminhos para aumentar a taxa a dezenas ou centenas de Hz (Monroe, Christopher, 2014).

A IonQ vem construindo gerações de trap chips e incrementando número de qubits. Segundo divulgações, o plano a médio prazo é concatenar vários chips em um mesmo aparato. Em 2021, a empresa anunciou estar desenvolvendo um sistema de múltiplos núcleos (chains) interligados; embora detalhes técnicos sejam escassos por serem propriedade, é sabido que investigações em curso envolvem fibras de poucas modas integradas aos chips de armadilha para coletar fótons sem ótica externa. Esse grau de integração reduziria perdas e permitiria taxas de entanglement remoto bem mais altas (talvez na faixa de kHz). O objetivo final declarado pela IonQ é escalar para centenas de qubits físicos ainda nesta década, e usar códigos de correção de erros distribuídos entre módulos para sustentar qubits lógicos.

Um diferencial da IonQ é a ênfase em all-to-all connectivity: mesmo em rede, eles almejam que qualquer qubit do sistema total possa interagir logicamente com qualquer outro, através de uma sequência de teleporte de portas. Isso contrasta, por exemplo, com arquiteturas que aceitam vizinhanças limitadas. A penalidade é exigir uma rede óptica de maior grau. Entretanto, Monroe et al. argumentam que comutadores ópticos (optical switches) de alta velocidade podem permitir reconfigurar conexões entre módulos rapidamente, efetivamente atingindo conectividade global sem saturar com hardware (Li, L. et al., 2015). Imagina-se, por exemplo, uma malha de beam splitters 2x2 que conecte dinamicamente pares de módulos sob demanda (um switch 8x8 poderia conectar 8 módulos em quaisquer combinações 4 pares simultâneos, por exemplo).

Em suma, o modelo IonQ de arquitetura modular baseia-se fortemente nos princípios discutidos: módulos homogêneos de íons (excelentes memórias e portadores de processamento local) + rede fotônica para entrelaçamento remoto + teleporte de portas para computação distribuída. Embora a IonQ não tenha ainda exibido publicamente um sistema multi-módulo completo, os componentes individuais e a pesquisa subjacente – muito creditada aos trabalhos de Monroe, Kim e colaboradores (Monroe et al., 2014) – indicam um caminho concreto. Com a maturação de tecnologias de fotônica integrada e avanços incrementais em velocidade de gate, é esperado que a IonQ apresente um protótipo de arquitetura modular de íons nos próximos anos, possivelmente conectando dois chips de 64 qubits cada, por exemplo.

### 3.4.2 Quantinuum (Honeywell) – QCCD e Teleporte de Portas em Múltiplos Módulos

A Honeywell Quantum Solutions (agora parte da Quantinuum, fusão com Cambridge Quantum) seguiu rota ligeiramente distinta da IonQ no início: priorizou a implementação completa do QCCD em um único chip antes de distribuir entre vários chips. Como discutido, o QCCD (Quantum CCD) é por si só uma abordagem modular, porém intra-chip: envolve zonas de armadilha (que podemos chamar sub-módulos) conectadas por movimentos de íons. Em 2021, a Honeywell demonstrou um sistema de 10 qubits aprisionados que integrava todos os ingredientes do QCCD – incluindo intersecções 2D para mover íons em diferentes eixos, resfriamento em background (usando espécies refrigeradoras), e uso de medidas intermediárias no circuito para teletransportar portas (Pino, J. M. et al., 2021). Em particular, reportaram a realização de uma porta CNOT teletransportada entre íons que não compartilham o mesmo conjunto de modos – ou seja, efetivamente entre zonas distintas da armadilha – utilizando medições e feed-forward clássico (Pino, J. M. et al., 2021). Essa porta teleportada teve

fidelidade elevada (~93%) e foi empregada para medir um quantum volume de 64, comprovando que o overhead do protocolo não impediu desempenho.

Tendo provado o QCCD num chip, o próximo passo lógico da Quantinuum é conectar múltiplos chips semelhantes. Eles mencionaram publicamente intenção de escalar adicionando mais “módulos QCCD” ligados por óptica. Uma possível configuração seria ter dois chips de armadilha separados (cada qual com 20 qubits, por ex.), com fibras permitindo entrelaçamento de um íon de comunicação de um chip com outro do segundo chip. Isso, combinado com a técnica de portas teleportadas (dominada no chip único), resultaria numa rede de dois módulos. Essa extensão ainda não foi mostrada publicamente, mas não há impedimento fundamental – possivelmente está em desenvolvimento.

Vale destacar diferenças sutis de filosofia: - A IonQ aposta em interconexão probabilística pura e correção de erros para superar baixos rates. - A Quantinuum vislumbra usar shuttling ao máximo e comunicação óptica apenas quando realmente necessário, possivelmente mantendo módulos dentro do mesmo invólucro criogênico para curtas distâncias de fibra (minimizando perda).

Os dois podem convergir, claro, mas exemplificam abordagens complementares dentro dos íons.

Um avanço notável do experimento de Pino et al. (2021) foi mostrar que se pode integrar múltiplos componentes ópticos e eletrônicos em um único sistema mantendo erros baixos. Isso sugere que mais módulos mecânicos (como outro chip armadilha) podem ser adicionados sem comprometer a fidelidade, se bem projetado. Além disso, eles evidenciaram que portas teleportadas são viáveis e quase tão boas quanto portas diretas – que é fundamental para qualquer rede modular (já que toda porta inter-módulo é teleportada).

Um desafio para a Quantinuum será a mesma questão de acoplamento óptico eficiente. No chip QCCD demonstrado, usou-se óptica bulk (lentes grandes, etc.) para entrega de laser e coleta de fluorescência. Para conectar chips diferentes, será necessário acoplar fótons em fibras ou guias integrados. A empresa não divulgou detalhes, mas há pesquisa na direção de chip traps com guias on-chip (o grupo de Sandia e MIT Lincoln Lab têm trabalhos a respeito). Quantinuum pode se beneficiar de muito know-how interno da Honeywell em sistemas ópticos robustos (dado histórico em aviônicos e controle de precisão).

Em perspectiva, a Quantinuum argumenta que seu design facilita escalabilidade porque podem replicar o tile de QCCD (que se mostrou de alto desempenho) e conectá-los. Notavelmente, error correction do tipo surface code exige malhas 2D de qubits locais – o QCCD facilita isso dentro de um módulo. Conectar módulos expande a malha além dos limites físicos de um chip. Em um artigo recente (PRX Quantum 2022, Ryan-Anderson et al.) o grupo inclusive demonstrou um ciclo básico de correção de erros (código de repetição) usando teleportação entre zonas, prelúdio de códigos distribuídos. O passo adiante seria códigos em que diferentes partes residem em chips distintos – possivelmente alvo em 2-3 anos.

Em resumo, a arquitetura da Quantinuum combina dois níveis: - Intra-módulo: QCCD e total conectividade via shuttling + teleportação no chip. - Inter-módulo: Enlaces fotônicos heráldicos entre íons de comunicação de diferentes chips, seguido de teleportação de portas lógicas para efetivar operações entre chips.

Essa é possivelmente a arquitetura de íons aprisionados mais completa visada no momento. A escala ainda é modesta (10-20 qubits por chip), mas viável de crescer para 50 ou 100 por chip em futuro próximo e talvez 4 chips interconectados (como IonQ planeja). A vantagem da Quantinuum é ter demonstrado experimentalmente cada elemento do quebra-cabeças (exceto a conexão de chips, que se infere próxima).

### 3.4.3 Xanadu – Arquitetura Modular Fotônica (Aurora)

No domínio fotônico, a Xanadu se destaca por buscar um computador quântico universal baseado em fótons com arquiteturas inovadoras. Conforme mencionado, sua recente revelação do sistema Aurora enfatiza a modularidade. O Aurora



consiste de quatro racks interconectados por fibra óptica, operando a temperatura ambiente, com um total de 12 qubits lógicos fotônicos distribuídos em 35 chips e ~13 km de fibra dentro do aparelho (Xanadu, 2025). Esses números dão uma ideia concreta: cada rack possivelmente contém várias fontes de fótons em chips (gerando estados comprimidos ou single photons), e as 13 km de fibra servem para criar atrasos temporais e conexões entre racks, formando um grande cluster state distribuído.

Segundo a Xanadu, essa arquitetura de racks modulares é escalável quase indefinidamente, bastando adicionar mais racks e fibras para conectar novos módulos. Cristian Weedbrook (CEO) afirmou que fotônica “resolveu a escalabilidade” pois se pode conceber milhares de racks e milhões de qubits interconectados. Claro, o diabo está nos detalhes: perdas ópticas. A Xanadu reconhece que o foco agora é diminuir as perdas para atingir tolerância a falhas. Entretanto, a abordagem modular é claramente vantajosa para fotônica: diferentes racks podem executar diferentes partes do circuito óptico (por exemplo, um rack pode gerar estados cluster pequenos, outro rack os funde através de medições multipartidárias, etc.). A natureza paralela da óptica permite que vários modos sejam propagados simultaneamente nas fibras, sem interferirem (salvo quando desejado nos beamsplitters), de modo que adicionar racks aumenta linearmente a capacidade de modos simultâneos.

Uma analogia que Xanadu faz é pensar em datacenters quânticos fotônicos, onde cada rack é como um nó de computação e a rede de fibras entre racks é como o backplane de interconexão. A diferença fundamental de arquiteturas com matéria é que aqui os qubits são os próprios sinais viajando. Em vez de qubits “morando” em cada módulo com conexões pontuais, na Aurora os qubits (pulsos de luz) viajam por todos os racks conforme necessário. A modularidade entra na habilidade de se construir subsistemas complexos independentemente: por exemplo, um rack pode abrigar um loop de fibra de 1 km que implementa um certo atraso específico e atenuação controlada – isso seria difícil de ter dentro do mesmo chassi que lasers e detetores, então separa-se em racks.

O Aurora integra 35 chips fotônicos – possivelmente chips de interferômetros programáveis (tipo meshes de phase shifters) e chips de fonte (ressoadores paramétricos que produzem estados comprimidos). Interconectar 35 chips dentro de 4 racks exige muita fibra e alinhamento. A Xanadu destaca que usou componentes comerciais e técnicas estabelecidas para montar Aurora, sugerindo que o desafio de engenharia foi principalmente de integração e não de desenvolver novos dispositivos físicos (Xanadu, 2025). Isso corrobora o argumento de modularidade: aproveitar componentes prontos e escalá-los modularmente.

Uma característica importante de fotônica é que os módulos podem ser bastante isolados eletricamente – não há campos EM fortes entre racks, apenas sinais ópticos confinado nas fibras. Isso simplifica a coexistência de muitos módulos, comparado a ter 100 módulos de íons com lasers possivelmente interferindo. Assim, a escalabilidade modular fotônica enfrenta mais problemas de pura ótica (perda, ruído nos detectores) do que de interferência entre módulos.

Em termos de desempenho até agora, Borealis (o protótipo anterior) alcançou supremacia quântica em amostragem, mas não era universal nem fault-tolerant. O Aurora pretende ser universal (capaz de portais arbitrárias via cluster states medidos) e eventualmente tolerante a falhas se perdas caírem. Para isso, a modularidade possibilita encadear códigos de correção de erros fotônicos: por exemplo, usar um rack inteiro para gerar um grande estado redundante (um código bosônico ou GKP states), e outro para processá-lo.

Resumindo, a arquitetura Aurora da Xanadu demonstra: - Modularidade em nível de sistema completo (racks interligados). - Grande quantidade de interconexão óptica (13 km de fibra interna). - Ênfase em escalabilidade: projetada para expandir para muitos racks sem mudar a filosofia. - Desafios atuais centrados em reduzir perdas ópticas e melhorar fidelidades de porta medida.

Essa filosofia contrasta com IonQ/Quantinuum em que fidelidades já são altas mas escalabilidade é o desafio. Na Xanadu, escalabilidade estrutural está encaminhada, mas as fidelidades e correção de erros são o calcanhar de aquiles a

trabalhar (Xanadu, 2025).

Finalmente, vale citar outras propostas modulares notáveis: a PsiQuantum (startup fotônica) planeja usar milhares de chips fotônicos interligados por fibra para um computador tolerante a falhas; a IBM tem projetos conceituais de quantum data centers conectando diversos processadores supercondutores via links de micro-ondas/ópticos. Embora não detalhados aqui, reforçam que a ideia de modularidade permeia toda a indústria quântica.

### 3.4.4 Comparação entre Abordagens

Após examinar diversas plataformas e estratégias, convém compará-las frente aos principais critérios de uma arquitetura quântica modular:

- **Tecnologia Física:** Íons aprisionados e átomos neutros oferecem qubits altamente coerentes (tempo de coerência longo,  $>1$  s) e facilidade de entrelaçamento remoto via fótons ópticos. Supercondutores oferecem qubits de operação ultrarrápida, porém com coerência relativamente curta ( $\sim 100$   $\mu$ s), tornando desafiador esperar por entrelaçamento remoto lento; sua interconexão favorece distâncias curtas (mesmo cryostat) ou exige transdução. Fotônica oferece distribuição nativa de qubits mas sofre com perdas e requer medidas para interações, demandando qubits redundantes para error correction desde o início.
- **Conectividade Local vs Remota:** Em íons, localmente é tudo-conectado (via fonons) mas relativamente lento ( $\sim \mu$ s a ms por porta); remotamente, acesso a qualquer outro módulo via fotônica, porém latência alta (ms) e probabilidade  $<1$  por tentativa – mitigável por memória longa. Em supercondutor, localmente conectividade é de vizinhos próximos (arranjo 2D), mas portas são ultrarrápidas ( $\sim$ ns) e fidelíssimas; remotamente, sem repetidor ou conversor, alcance é limitado a cryostat local – há demonstração de 4 chips integrados em nm escala com ligação determinística  $\sim 100$  ns (Gold. A., et al, 2021), mas para  $>$ metro, precisaria converter para óptico com overhead. Em fotônica, conectividade "local" significa dentro de um chip óptico (vários modos interferindo) – que pode ser de grau alto mas limitado pelo número de guias; conectividade remota trivial pelo linking de fibras – pode ser total (qualquer chip a qualquer chip via fibra), o problema é perda incremental por distância.
- **Escalabilidade em Número de Qubits:** Supercondutor atualmente lidera em qubits físicos num chip monolítico (IBM e Google  $\sim 433$  qubits num dispositivo). Porém, cada qubit extra complica controle e aumenta cross-talk. A ideia modular no supercondutor é reduzir complexidade por chip: por ex, 4 chips de 128 qubits cada em vez de 1 chip de 512 – melhorar yield e reduzir cross-talk (Niu, Jingjing et al, 2023). Íons aprisionados e neutros tipicamente têm  $<100$  qubits por módulo hoje; modularidade permite ir a vários módulos e somar qubits linearly. A IonQ e Quantinuum planejam escalar nessa linha a centenas de qubits físicos agregados (mas efetivos lógicos ainda poucos sem QEC). Fotônica, em princ, escalabilidade de qubits entrelaçados é altíssima (Borealis demonstrou 216 modos) – mas nota: eram qubits muito ruidosos (perdas etc.). O que se nota: monolítico vs modular, cada abordagem tem doping: SC aumentou monoliticamente, mas agora terá que modular; íons já modulam cedo; fotônica modular intrínsec.
- **Tolerância a Falhas e Overhead:** Para atingir computação tolerante a falhas, estima-se necessários milhares de qubits físicos por qubit lógico. Em SC, dada alta fidelidade, esse overhead pode ser menor (talvez  $\sim 1000:1$  ou menos com melhorias); em íons, fidelidades também altas, overhead similar – porém latências maiores podem complicar ciclo QEC. Em fotônica, as perdas implicam overhead gigantesco (talvez  $1e6:1$ ) a menos de avanços. Comparando: SC tem caminho mais claro a fazer código de superfície e implementar repetitivamente (já testam pequenos patches); íons podem também, mas teleporte entre módulos não pode demorar mais que o tempo de

correção antes decoerir correções – isso requer entanglement rate decente ou usar qubits memória muito estáveis (nesse ponto íons ok). Fotônica requer implementar QEC inteiro via medidas – conceitualmente possível mas precisam compressão de perda via códigos bosônicos.

- Recursos de Engenharia: SC requer criogenia pesada e eletônica RF proximidade – modulação significa múltiplas cryos ou cryos maiores, e interconnects coax/opt complicados. Íons requerem lasers estabilizados e vácuo ultralto – replicar isso para muitos módulos é não trivial (pode exigir um setup de laser por módulo, ou fibra repassando lasers, etc). Átomos neutros similar a íons. Fotônica requer lasers e detectores, mas estes são todos em sala normal – lasers podem ser compartilhados, e detectores escalados duplicando racks.
- Maturidade Experimental: SC qubits modulares: mostrados 4 chips entangled (Gold. A., et al, 2021) e 2 chips entangled via coax (ETH 2018) – incipiente. Íons modulares: já mostrados entanglement 2 trap (NIST/U.Md, 2006-2015), e teleporte porta in chip (Quantinuum 2021) – progressos notáveis, mas nenhum produto multi-modular ainda. Átomos neutros modulares: experim. de entanglement de 2 arranjos de átomos (Harald Weinfurter 2020, entangling Rb atoms in separate labs). Fotônica modular: Xanadu Aurora 4 racks working (ainda não se sabe performance detalhado, mas lançado). No total, cada abordagem tá dando passos, ainda longe de escalas massivas.
- Vantagens Exclusivas: Íons e neutros podem aproveitar rede de quantum internet emergente – integrando-se a futuras redes criptografia e relógios, um plus. SC e fotônica se adequam a datacenters (SC por integração com controle clássico avançado, fotônica por ambient). Fotônica e neutros escalam geograficamente para redes sem esforço (dá para ter módulos em cidades diferentes se quiser). SC e íons preferem proximidade por coeren.

Em linhas gerais, não há “vencedor absoluto” – cada abordagem modulare tem trade-offs. Íons aprisionados talvez estejam numa posição doce: escalabilidade modular está incrustada em seus planos desde cedo e coesão do sistema pode manter fidelidades altas, a custo de menores clocks. Supercondutores precisaram vencer na marra problemas monolíticos e agora buscam modularidade para pular de 1000 para  $1e5$  qubits – eles têm o desafio extra de transdução se quiserem ir além de um refrigerador. Fotônica é possivelmente a mais escalável topologicamente, mas necessita avanços em diminuir perdas para ser computacionalmente útil.

Uma comparação sumarizada é apresentada a seguir, na Tabela 1:

**Tabela 1:** Comparativo dos parâmetros de conectividade.

<b>Critério</b>	<b>Íons Aprisionados</b>	<b>Supercondutores</b>	<b>Átomos Neutros</b>	<b>Fotônicos</b>
<b>Conectividade local</b>	Total dentro módulo (via fonons), ~µs gates	Parcial (vizinhos em malha 2D), ~ns gates	Parcial (arranjo geométrico, via Rydberg)	N/A (fótons passam por circuitos ópticos)
<b>Conexão entre módulos</b>	Fotônica (fibras, ent. heráldico)	Coax/Resonador (curto alcance); opt. transd.	Fotônica (fibras + cavidades p/ boost)	Fibra óptica (fóton naturalmente viaja)
<b>Latência remota</b>	ms (taxa ~100 Hz), memórias robustas (s)	~ns-µs (se chips no mesmo cryo); long dist dif.	ms (com cavidade, poss vel MHz rates)	~ do tempo de voo de luz + detecção (µs-ms)
<b>Fidelidade local</b>	99.9% 2-qubit	99.9% 2-qubit	~99% (Rydberg gate)	Medida-projetiva tip. 90-99% (dep det etc.)
<b>Fidelidade remota</b>	90-95% demonstrada p/ 1 par (pode purificar)	99% (4-chip Gold determinístico)	80-90% (exp recentes)	? (depende de perda; Borealis cluster quality ~ baixa)
<b>Qubits/módulo</b>	10-50 (demonst.), visionado 100+	100-1000 (IBM condor chip 433, mod 4 chip ~32 each)	50-300 (arranjos dem.)	3-12 lógicos (Xanadu Aurora), potencial físico >> (muitos modos)
<b>Desafio maior</b>	Aumentar entanglement rate entre módulos	Integrar muitos chips e conversão p/ long dist	Perdas e acoplamento a fibras nanocavidades	Perdas ópticas, exigência de códigos QEC overhead altíssimo
<b>Estado da arte</b>	Teleporte CNOT entre zonas (Quantinuum 2021)	4 chips entangled (npjQI 2021), 127q monolit IBM	2 átomos ent. 33km (2021), 256 at arranjo monolit	Supremacia com cluster 216 (Xanadu 2022); Aurora 2025 mod 4 racks

Fonte: Dados da Pesquisa (2025).

Cada abordagem possivelmente convergirá para um ponto onde erros e latências são equivalentes à custo computacional comparável – ou seja, no limite todas precisariam implementações de QEC. Nesse regime, importará mais a contagem de qubits físicos disponível e a taxa de porta (throughput). Pode ser que supercondutores e fotônica sobressaiam em throughput (muitas gates/s), enquanto íons/neutros em qubits de alta qualidade, mas gates lentos. A melhor arquitetura final pode até ser híbrida (e.g., usar íons como memórias e SC como processadores interligados, etc.), mas isso está além do estado atual.

Considerações de uso: Se o objetivo é rede quântica distribuída (quantum internet), átomos e íons integrados com fotônica são naturais – Google e Quantinuum planejam até satélites com íons. Para computação intensiva num local, SC ou fotônica poderiam ser preferíveis por taxa. Isso ilustra que "melhor" depende do aplicativo. Modularidade permite inclusive que diferentes módulos tenham especializações – ex: um módulo sensor, outro computa etc.

#### 4. Conclusão

Arquiteturas modulares e redes quânticas representam a fronteira para escalar a computação quântica além das limitações dos dispositivos isolados. Neste artigo, analisamos os princípios dessas arquiteturas, as plataformas físicas envolvidas e exemplos concretos de implementação. Argumentamos que, tal como na evolução dos computadores clássicos, a modularidade será chave para construir computadores quânticos com milhares ou milhões de qubits.

Os **conceitos fundamentais** de arquiteturas modulares quânticas residem em dividir um grande processador em unidades menores (módulos) interligadas por canais quânticos capazes de distribuir entrelaçamento. Essa abordagem mitiga problemas de controle e decoerência que afetam sistemas monolíticos muito grandes, ao custo de introduzir novos desafios: latência de comunicação, necessidade de entrelaçamento remoto confiável e protocolos distribuídos de computação.

Examinamos os **tipos de hardware** que habilitam arquiteturas modulares: qubits supercondutores, íons aprisionados, átomos neutros e qubits fotônicos. Cada tecnologia oferece um balanço distinto entre velocidade, fidelidade e alcance de interconexão. Em particular, íons e átomos, com suas transições ópticas adequadas, se acoplam naturalmente a redes fotônicas, enquanto supercondutores demandam interfaces de conversão para conectar módulos distantes. Já qubits fotônicos são por definição nativos de redes, e seu desafio reside em implementar processamento lógico robusto dado seu caráter de voo livre.

Discutimos as **conexões entre módulos** – tipicamente feitas por entrelaçamento fotônico – e limitações físicas como perdas e decoerência. Destacamos que estratégias de **redes quânticas** como repetidores e teleporte de portas lógicas são essenciais para escalar distâncias e efetuar computação distribuída. Apresentamos modelos de **custo e latência** que quantificam o impacto de operações remotas, reforçando a importância de maximizar paralelismo e minimizar comunicações desnecessárias em algoritmos distribuídos.

Através de **estudos de caso** (IonQ, Quantinuum, Xanadu), vimos diferentes implementações do conceito modular: desde íons aprisionados conectados por fibra, passando por múltiplos chips supercondutores acoplados deterministicamente, até racks fotônicos modulados por extensas linhas de fibra. Essas iniciativas demonstram que a modularidade não é apenas teoria – protótipos reais já exibem elementos modulares e confirmam a viabilidade de se construir máquinas quânticas maiores combinando subsistemas.

Na **comparação das abordagens**, concluímos que nenhuma plataforma resolve todos os desafios sozinha. Cada qual aproveita seus pontos fortes – fidelidade, velocidade ou conectividade – e lida com seus pontos fracos – sejam perdas ópticas, decoerência ou complexidade de controle. A **escolha da arquitetura modular ótima** poderá depender da aplicação: redes quânticas de longa distância favorecem átomos/íons, enquanto computação local intensiva pode favorecer supercondutores ou fotônica integrada. É possível que arquiteturas híbridas emergjam combinando tecnologias para aproveitar sinergias.

O caminho à frente apresenta obstáculos consideráveis. Será necessário aprimorar tecnologias de interface quântica (transdutores quânticos, cavidades integradas), desenvolver protocolos de correção de erros capazes de tolerar enlaces ruidosos (Nickerson, Fitzsimons & Benjamin, 2014), e aperfeiçoar a engenharia de sistemas para permitir integração de dezenas de módulos complexos. A pesquisa atual, porém, indica um progresso contínuo: taxas de entrelaçamento remoto melhorando, demonstrações de teleporte de operações multi-módulo (Pino, J. M. et al., 2021), e expansões das capacidades de controle e fabricação de qubits em escala.

Em última instância, a modularidade pode se tornar transparente para o usuário, tal como em computadores clássicos atuais ninguém percebe que há múltiplos núcleos ou vários processadores trabalhando em paralelo. Para chegar lá, a comunidade científica e de engenharia deverá tratar o sistema quântico distribuído como uma entidade coesa – um verdadeiro **computador quântico em rede**. Quando isso for alcançado, teremos dado um passo crucial rumo à realização de computadores quânticos de grande escala, possivelmente capazes de solucionar problemas hoje inabordáveis.

## Referências

- Awschalom, D. et al. (2021). *Quantum interconnects for the long-distance quantum internet*. PRX Quantum. 2, 017002.
- Brown, K. R., Kim, J. & Monroe, C. (2016). Co-designing a scalable quantum computer with trapped atomic ions. *NPJ Quantum Information*. 2, 16034.
- Bushman, J. (2025). *Parallel atom-photon entanglement paves way for future quantum networking*. The Grainger College of Engineering (University of Illinois Urbana-Champaign). <https://grainger.illinois.edu/news/stories/77824>.
- Cantori, S., Pfaffhauser, M., Bäumer, E., Scafirimuto, F. & Davis, R. (2025). *Using dynamic circuits to efficiently implement quantum states with long-range entanglement*. IBM Quantum Computing Blog. <https://www.ibm.com/quantum/blog/long-range-entanglement>.
- Chen, Y. et al. (2021). (Google Quantum AI). Exponential suppression of bit or phase errors with cyclic error correction. *Nature*. 595, 383–7.
- Covey, J. P., Weinfurter, H. & Bernien, H. (2023). Quantum networks with neutral atom processing nodes. *NPJ Quantum Information*. 9(90).
- Gold, A. et al. (2021). Entanglement across separate silicon dies in a modular superconducting qubit device. *NPJ Quantum Information*. 7(1), 142.
- Hucul, D. et al. (2015). Modular entanglement of atomic qubits using both photons and phonons. *Nature Physics*. 11, 37-42
- Krut'yanskiy, V. et al. (2023). Entanglement of trapped-ion qubits separated by 230 meters. *Phys. Rev. Lett.* 130, 050803.
- Li, X-G. et al. (2025). Direct evidence for cosmic-ray-induced correlated errors in superconducting qubit array. *Nature Communications*. 16, 4677.
- Madsen, L. S. et al. (2022). Quantum computational advantage with a programmable photonic processor. *Nature*. 606(7912), 75-81.
- Matsukevich, D. N. et al. (2006). Distribution of entanglement using photons and phonons in a crystal network. *Science*. 311, 1008-12.
- Monroe, C. et al. (2014). Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects. *Phys. Rev. A*. 89, 022317.
- Monroe, C. (2014). *Modular Entanglement of Atoms*. Quantum Computing with Trapped Ions. <https://iontrap.duke.edu/2014/11/17/modular-entanglement-of-atoms/>.
- Nickerson, N. H., Fitzsimons, J. F. & Benjamin, S. C. (2014). Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. *Phys. Rev. X*. 4, 041041.
- Niu, J. et al. (2023). Low-loss interconnects for modular superconducting quantum processors. *Nature Electronics*. 6(3), 235-41.
- Pereira, A. S., Shitsuka, D. M., Parreira, F. J. & Shitsuka, R. (2018). *Metodologia da Pesquisa Científica*. Santa Maria: Editora da UFSM
- Pino, J. M. et al. (2021). Demonstration of the trapped-ion quantum CCD computer architecture. *Nature*. 592, 209–13.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*. 104, 333–9.
- Rother, E. T. (2007). Revisão sistemática versus revisão narrativa. *Acta Paulista de Enfermagem*, São Paulo. 20(2), v–vi.
- Sunami, S. et al. (2025). Scalable networking of neutral-atom qubits: Nanofiber-based approach for multiprocessor fault-tolerant quantum computers. *PRX Quantum*. 6, 010101.
- Xanadu (C. Weedbrook). (2025). Press Release: *Xanadu announces Aurora, a universal photonic quantum computer with a modular architecture*. The Quantum Insider.