Quantum Memories to Grow the Quantum Internet

contact: Michele Reilly CTO, Turing Inc
mmreilly@mit.edu | mr@turingquantum.com

July 15th 2017

Contents

1 Executive Summary

In this white paper, we provide an overview of the technology behind the Turing QuSTICK: A quantum communications device for secure networking that is modular, scalable, portable, and can be deployed anywhere.

Turing's technology is not limited to line-of-sight communications as is typical with today's quantum communications methods. The QuSTICK is breakthrough hardware that is technologically and creatively superior to existing attempts – attempts that at present are failing in the marketplace due to the cumbersome and inadequate nature of the infrastructure.

The Turing QuSTICK technology addresses the vulnerabilities of public-key cryptography, classical key-exchange, and so-called trusted couriers. The intelligence community and other private security companies are still too reliant on the fragile human factors that often fail – complex webs of trust and interdependent hierarchies of people. There is a pressing need for truly secure communication, both on the consumer side and for military applications.

The contemporary security paradigm is a faith-based one: it relies on the fact that quantum computers will never be built. However, we know efforts to build quantum computers are underway today. This implies an unmet urgent need for businesses and governments to make investments now in order to be secure against attacks in the future. Banks, medical companies, sovereigns, and corporations with critical data to communicate and store do not have the luxury of waiting for the first attack to address their vulnerabilities.

With our communications technology, we have tackled four main issues:

- **RANGE:** Quantum communications protocols such as Quantum Key Distribution (QKD) need to operate over intercontinental distances. Networks that are limited to a few kilometers do not have significant practical impact to motivate large-scale investment. Most of the important applications require the ability to communicate using quantum mechanics at global scales.
- **ACCURACY:** Accuracy refers to the errors associated with a quantum communications channel, commonly referred to as Fidelity. The communications link for QKD, or any other quantum protocol, requires extremely high Fidelity (low error rates) for transmission over whatever target distance is desired.
- **RATE:** In today's data-hungry world, any communications process working at low data rate, such as a few hertz, is useless – whether it is quantum or not. We need high data rates in any quantum protocol to ensure its usefulness with the high bandwidth classical infrastructure we already have available.
- **COST:** Infrastructure and maintenance costs are always of significant concern with any technology. Only in very rare cases are we willing to accept a technological infrastructure base that costs billions or trillions of dollars to build and maintain. The same is true for quantum technology. Something that is extremely expensive to build, deploy and maintain will only be acceptable if it offers a commercial or strategic advantage of comparable value.

Each of these considerations must be addressed in order to deliver a robust and truly secure communications system. *We are the only company targeting all four from the start.* For instance, many of the techniques currently under development for realizing Quantum Key Distribution (QKD) links are limited to short distances and have errors in the 10-20% range. They can barely reach 10-100 Hz at their maximum range, and will be very expensive to deploy. Quantum satellites [TCT+16, TCCF+17, YCL+17] and quantum repeaters [FWH+10, SSdRG11, MSD+12, ATL15] are examples these very limited technologies.

The discrepancy between where such systems *can* be deployed and where we *want* them deployed is the primary reason why few of the first businesses dedicated to QKD technology have not seen

significant growth, even though they have been in existence for well over a decade. Simply put, these companies are selling technology that nobody wants to buy because they cannot be used in the arenas where they matter the most.

Being able to build a quantum key-distribution (QKD) system does not make economic sense for connecting a server room to another server room down the hall, or across a military base, as many other security techniques are feasible at that scale and are simply more cost-effective to implement.

1.1 Where will Turing's technology be valuable?

We can deploy secure key distribution technology to field units around the world, even those in hostile or inhospitable environments. There is no quantum communications technology in existence today other than our QuSTICK that makes this possible. It is light and portable. We are able to scale smoothly from key distribution to real time secure data communication, and ultimately, to a Quantum Internet.

To fully address the question of where Turing's technology will be valuable, we will take a look at the practicalities of existing communications security practices and where the major points of vulnerability lie.

1.1.1 Key Exchange vs. Data Exchange

Quantum Key Distribution (QKD) is one aspect of quantum cryptography. QKD refers to the distribution of encryption keys. Such keys are used to encrypt and decrypt data that is then transmitted via conventional methods.

Key material is typically *significantly* smaller than the encrypted message itself. To transmit classical information using a quantum network directly will take significantly more physical resources than simply exchanging an encryption key (which may only be on the order of a thousand physical bits). The security of quantum key distribution can be proven mathematically without imposing any restrictions on the abilities of an eavesdropper – something not possible with classical key distribution. This is usually described as "unconditional security". Ultimately, the underlying technology of QuSTICK can be scaled to support quantum encryption of entire messages.

1.1.2 Public-Key Cryptography

Public-key is a method of encryption that utilizes an asymmetric process based on the assumed difficulty of certain mathematical problems. Probably the most well-known protocol, RSA, is based on the difficulty of factoring large composite numbers into their prime components.

Thus far, public-key cryptography has seen its most successful adoption in the online space, where cryptographic security is needed for many sensitive tasks: online banking, email and messaging. But exchanging secure cryptographic keys is a relatively slow process, which makes it impractical for many important applications (for example, a website with millions of clients who need services provided in real-time).

Public-key cryptography uses two keys that are related through a asymmetric mathematical function. A asymmetric function in this context is a problem that is very easy to solve in one direction, such as multiplying two large prime numbers together, but extremely difficult to do in reverse (finding the prime factors of a large composite number). One of the two cryptographic keys is published openly and used by a sender to encrypt a message. The other key is held in secret by the receiver and is needed to decrypt the message. Unless an adversary has the ability to solve the difficult underlying mathematical problem, the message cannot be decrypted without the private-key.

While public-key encryption has been widely adopted across the world, it is not considered information-theoretically secure due to the *assumed* difficulty of the underlying mathematical problem. In the case of RSA and prime factorization, we don't fully understand the computational difficulty of the factoring problem, and can therefore not guarantee that RSA will not be broken tomorrow by some brilliant computer scientist or mathematician.

Quantum computing adds a whole new level of security problems to public-key cryptography. It has been shown that, if an adversary has access to a large quantum computer, the mathematical problems underlying public-key cryptography *can be solved efficiently.* Shor's algorithm, the most famous quantum algorithm, kick-started the first acceleration of quantum computing development in 1994. Dr. Shor demonstrated that both the factoring and discrete log problems could be solved using a polynomially increasing number of quantum bits (qubits) as the problem size increases. This effectively makes a large number of public-key crypto-protocols obsolete if quantum computers could ever be built. Consequently, *any* sensitive information transmitted using public-key cryptosystems has to be assumed compromised by the most hardened security experts, even though quantum computers capable of breaking public-key cryptosystems (presumably) do not yet exist.

1.1.3 The Problem of Long-Term Storage

The problem is more pronounced with information that requires long-term security after its initial transmission. This may include national security secrets, trade or business secrets or personal information such as medical and banking records. The risk when transmitting this kind of information is that its value may not be related to how quickly it can be decrypted, but instead whether it can be decrypted at all. It is well known that intelligence operations (both by governments, private sector actors or nefarious organizations) routinely intercept encrypted traffic on classical networks and simply put it into long-term storage, hoping that, some day in the future, technology progresses to the point where the information can be decrypted and exploited.

It is this kind of information that is particularly vulnerable to quantum computers. If data is encrypted using protocols that are susceptible to quantum attacks, eventually this information will become exposed. If secrecy needs to be maintained many years in the future (which is certainly the case for much national security, trade and business secrets and personal information), this is a significant problem – one that does not yet have a satisfactory solution beyond either preventing the encrypted information from being intercepted in the first place or crossing our fingers that quantum technology will either not be built or only be available to forever friends and allies.

One phrase that floats around the community more and more is "post-quantum cryptography". To be clear, post-quantum cryptography is neither "post", nor "quantum", nor "cryptography". It is a term used to describe any classical cryptographic technique that is based on mathematical problems that can be proved to be *not efficiently solvable by quantum computers.* While research is still ongoing, there have been no significant results suggesting such a crypto scheme exists. Schemes such as lattice-based cryptography and other symmetric schemes have not yet been proven crackable by a quantum computer, but complexity classifications for such schemes have not in general been proven for either the classical or quantum space. Consequently, utilization of such schemes cannot be guaranteed to be secure into the future.

1.2 One-time pads and key exchange over a network

The concept of informationally secure cryptographic schemes dates back to the late 19th century. Initially developed by Frank Miller in 1882 and re-invented in 1917, it wasn't until the 1940's that Claude Shannon actually theoretically proved the power of this technique to completely secure a transmitted message. To this day, one-time pads remain the only officially provable informationally secure protocol for cryptography.

Such a protocol is a method to encrypt information such that, *if implemented perfectly* (and this is where we actually get into trouble), can never be broken. The laws of the universe and reality itself guarantee that an adversary without access to the encryption key *can never*, regardless of any hypothetical computational power (quantum, classical or anything else) decrypt the message.

Informationally secure protocols are the holy grail of network and information security and are of huge relevance to almost all aspects of digital life in the twenty-first century. Unfortunately, as with most things in the real world, the devil is in the details: The caveat of *perfectly implemented* is practically impossible to achieve.

The basic principles and core assumptions of one-time-pad encryption are:

- The sender (conventionally referred to as Alice) and receiver (Bob) both have access to a shared key that is completely random. Here we mean "random" in the information-theoretic sense of the word in that the bit strings have maximum Shannon entropy. This key has exactly same length as the message to be transmitted.
- The key that Alice and Bob share cannot have been compromised in any way, i.e. no adversary can have any information about any part of the random bit string shared between Alice and Bob at any time.
- This key is used once *and only once* to encrypt a single message. The key or any part of the key can never be used to encrypt two separate transmissions.

If these three core points are adhered to, the encrypted message is completely unbreakable by any adversary who may intercept it, regardless of any hypothetical computational resources they may have. Unfortunately, the above conditions are simply not practical in most real world situations.

- Generating a random bit string, in the true definition of random, has only become technologically possible in the past decade or two as result of quantum technology. Quantum random number generators use the intrinsic randomness of quantum mechanical processes to generate random bit strings that have maximum entropy. These devices can now be bought as off the shelf units.
- Avoiding the reuse of a key can, in principle, be adhered to; but satisfying this requirement may be economically or logistically impractical. Avoiding reuse becomes even more difficult when the message to be encrypted is big. There are many examples of technology in the field that require secure, high data-rate transmission. Video footage from helicopters or drones is a good example. A drone equipped with a high-resolution 4K camera system needs to be able to transmit information at a rate of 764 GB/hour uncompressed. This would require the same rate of one-time-pad material to be generated and shared between Alice and Bob to ensure a secure connection of the video stream. None of this material could be used even a second time.
- The third and most vulnerable constraint is the sharing of key information between Alice and Bob in a manner that can be guaranteed to have never been compromised. This is complicated by the fact that Alice and Bob in almost all practical situations will not be in physical contact with each other prior to sharing an encrypted message. How can we exchange a large amount of key material between two parties who have never met in such a way that we can be as certain as humanly possible that the key information was not revealed, sold, copied or otherwise accessed by a third party? The third party may even be authorized to have access to the key information. This would immediately destroy the *perfectly implemented* assumption underlying the security of the one-time-pad.

While many crypto-systems are not based on one-time-pad encryption because of these constraints, the sharing of cryptographic keys is still of major concern. Transmitting keys over *any* publicly accessible network is fraught with security problems as there is no way of encrypting the key material without then asking how the keys used to encrypt the keys are shared.

For extremely sensitive tasks, key-exchange and security is usually achieved through some combination of segregated networking and obfuscation. Here are some examples:

1.2.1 Security through obfuscation

Some techniques for key and message exchange attempt to hide from adversaries any evidence that such protocols are taking place. This may be as simple as giving a secret key or message to a

20-something hipster riding a bike instead of man in a well-tailored suit with dark sunglasses and a briefcase handcuffed to his wrist. Security through obfuscation is more focused on not arousing suspicion or misdirecting adversaries to use a false attack vector when attempting to compromise information exchange, rather than directly defending against a potential attack.

1.2.2 Security through segregated networks

Segregating classical networks from the public and/or potential adversaries is a more common technique simply because the speed and range of radio signals and optic fiber technology is so large. Ensuring that a military or intelligence network is not connected to anything that is more widely accessible, in principle, allows for a higher level of security and the ability to have confidence that cryptographic keys can be sent without interception.

These basic techniques are rarely implemented in isolation. Complex network security uses many combinations of good encryption, segregated networks, obfuscation and other techniques to lower the probability of compromised data transmission. However, none of these techniques fully satisfy the assumptions of an information-theoretically secure cryptographic protocol, and they often rely heavily on the honesty and competence of human personnel in order to be implemented effectively.

1.3 QKD versus classical key exchange techniques

A 2016 assessment from the UK agency GCHQ https://www.ncsc.gov.uk/whitepaper/quantum-key-distribution references four concerns with QKD (Quantum Key Distribution) as an effective replacement for classical techniques:.

These concerns are:

- *QKD protocols address only the problem of agreeing on keys for encrypting data.* Ubiquitous on-demand modern services (such as verifying identities and data integrity, establishing network sessions, providing access control, and automatic software updates) rely more on authentication and integrity mechanisms — such as digital signatures — than on encryption.
- *The two major functional limitations of commercial QKD systems are the relatively short effective range of transmission and the fact that BB84 and similar proposals are fundamentally point-to-point protocols.* This means that QKD does not integrate easily with the Internet or with the mobile technologies, apps and services that dominate public and business life today.
- *Hardware is relatively expensive to obtain and maintain.* Unlike software, hardware cannot be patched remotely or cheaply when it degrades or when vulnerabilities are discovered.
- *Any real-world QKD system will be built from classical components, such as sources, detectors, fibers, and ancillary classical network devices, any one of which may prove to be a weak link.* A number of attacks have been proposed and demonstrated on deployed QKD systems that subvert one of more of these hardware components, enabling the secret shared key to be recovered without triggering an alarm.
- *Denial of service (DoS) attacks that interfere with the paths carrying the QKD transmissions also seem potentially easier with QKD than with contemporary Internet or mobile network technologies.* Since QKD devices typically abort a key establishment session when they detect tampering, this makes it difficult to recommend QKD for contexts where DoS attacks are likely to be attempted.

As we describe the Turing system, we will address each of these concerns and examine how it can resolve or significantly mitigate the chief concerns of a quantum-based key distribution network.

2 The Turing QuNET

The Turing QuNET is a combined hardware/software protocol stack that is designed specifically to shrink the vulnerability window of key exchange to close to nothing. Our technology replaces many of the human-centered aspects of message encryption and gets as close to satisfying the assumptions of an informationally-secure cryptographic protocol as possible. We leverage active portable quantum hardware and newly invented quantum authentication and networking protocols to design a networking structure that can be deployed in many different arenas to:

- Provide quantum authentication tokens that can be used to identify personnel and equipment.
- Exchange cryptographic keys, including one-time-pad material at a high data rate.
- Transmit information directly over quantum channels.
- Couple into future public quantum communication networks without the need for segregation.

The Turing QuNET system is what we call an *scalable quantum network*, where new protocols come online as we produce and distribute more unit volume. Only a few qubits are required for protocols such as authentication tokens; and, as we build more and more of the fundamental hardware, the QuNET begins to scale to support key exchange, one-time-pads, direct data transmission and ultimately the quantum version of the Internet. This expansion does not require any re-design of the underlying hardware or network protocols.

All protocols can be run simultaneously across the network without the need for separate protocols, network segregation or hardware. Network capacity and protocol flexibility becomes purely a function of the number of QuSTICK units deployed in the field. This is quantitatively described later in this document once we detail the estimates needed for each application stage of the QuNET in Fig. 13

The other significant aspect of the Turing QuNET is its practical flexibility. As we will describe, the QuNET system can be deployed anywhere sensitive data exchange is needed. This includes, but is not limited to:

- Large military bases, intelligence or governmental installations.
- Mobile platforms such as aircraft, naval ships (including submarine assets) and satellites.
- Field deployment, with mobile units carried by individual officers or diplomatic staff.
- Highly sensitive assets such as long range missiles

QuNET nodes (which we name Turing QuSTICKs) are portable, active quantum technology units that connect to each other through a quantum entanglement network that cannot be directly disrupted or hacked. The portability of QuSTICK units allow us to deploy nodes of the QuNET essentially anywhere on earth and the nature of our protocols allows us to use this network for both cryptographic key exchange and data transmission in a manner that is vastly superior to any current software techniques.

As we will describe, the QuNET system also has the ability to evolve and be incorporated into a future commercial quantum communications network without suffering any loss of information security, allowing for classified and unclassified data to exist on the same hardware infrastructure.

2.1 Trusted couriers and one-time-pads

Before detailing the specifics of the Turing QuNET, let us first discuss the general framework that our system leverages. This is the principle of physical trusted couriers [Mer78] and one-time-pads [Sha49]. This general technique, utilized routinely for extremely sensitive key exchange, does not transmit key material over any classical network and instead entrusts a physical courier with pre-loaded key material on a portable memory stick.

The key material, which is generated at home base, is physically couriered to its destination and security is, in principle, achieved using a combination of obfuscation, deception and trust of

the courier. The couriers themselves can be heavily vetted, achieving the highest levels of security clearance within an organization and dummy key material may be given to couriers at random to see if it is showing up in places where it is not supposed to.

Today's technology makes the other requirements – easy access to high capacity storage and one-time-pad material of sufficient quantity to encrypt the largest-data-rate applications – straightforward to achieve. Consequently, distribution of the key material itself is typically the hardest assumption to satisfy when relying on the information-secure nature of one-time-pads.

The Turing QuNET is a technological solution to the problem of having to trust the courier. We exploit the nature of quantum to develop an analogue of the trusted-courier network wherein the human couriers are no longer a weak link.

3 Turing Protocols

There are several protocols that can be sequentially realized using the QuNET framework, with each new protocol becoming possible as we increase the number of QuSTICKS manufactured and deployed in the field. We can specify five layers of cryptographic related protocols as a function of the number of QuSTICKS available to the QuNET, as illustrated in Figure 1:

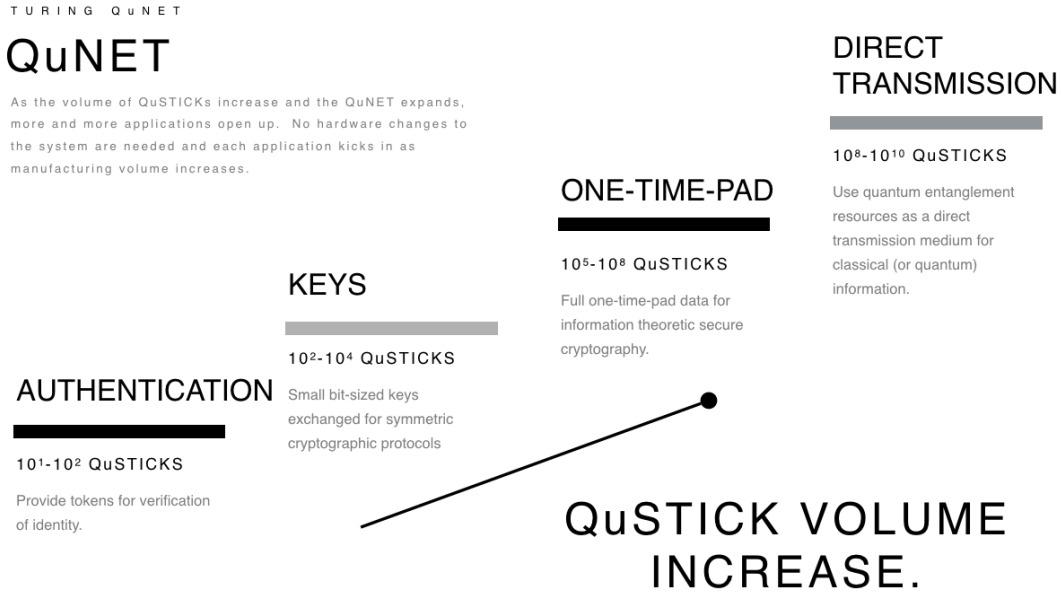

Figure 1: Developing a quantum communication system that is modular allows the system to naturally expand as more units become available. The building block of our system is the QuSTICK. As more and more units are manufactured and deployed, applications "come online". Units are not replaced, but rather add to the capacity of the preexisting network.

Each of the first three protocols – authentication, keys and one-time-pads – rely on the same basic principle: the ability for two users to share remotely a finite amount of correlated random information. The use of this data for each individual protocol becomes a function of the *amount* of the material you have. Authentication keys or tokens require only a few bits, cryptographic keys require a larger number of random bits, and one-time-pads an even larger amount (in principle, the size of the message to be transmitted). The final application in the QuNET stack, namely direct transmission of classical information through a quantum network, is different in nature and requires the largest volume of quantum devices to realize.

One element of the QuNET stack outside the scope of this white paper is the direct transmission of a large amount of quantum information. This would be used, for example, in a Quantum

Internet, where actual quantum computing hardware is connected together in an analogous manner to the classical Internet. The Turing QuNET does have the ability to be deployed in this fashion once production volumes reach appropriate levels. The utilization of the Turing network for quantum computation and communication systems will be discussed in a separate document.

4 The Turing Chip Set

The key innovation for the Turing QuNET is the fabrication and packaging of a long-lived *active* quantum memory. Quantum memories [Ter15] are inherently fragile due to quantum decoherence induced via uncontrollable interactions with the wider environment. Today's best physical qubits typically have lifetimes of the order of one millisecond. While there has been significant research into finding or engineering quantum systems with lifetimes of seconds or minutes [ZHA+15, RHAS17, AGA+18], these systems are effectively uncontrollable because they have been engineered to be so isolated from the environment that the control signals required to load, manipulate or measure quantum information are impossible to implement. The ability to build quantum technology that is useful in the field requires increasing the effective memory time – the lifetime of data stored as qubits – so that they can be used on timescales relevant to humanity.

The problem of increasing the effective lifetime of qubit information was solved with the formulation of active quantum error correction (QEC) protocols [DMN13]. QEC utilizes redundant encoding, wherein a single *logical* qubit of quantum information is encoded using a finite number of *physical* qubits. Continuous operations on this array of physical qubits are used to extract information regarding physical errors; and, provided physical error rates are sufficiently low, the logically encoded information can be preserved for an arbitrary amount of time. Unlike specially engineered quantum systems that are isolated to achieve long memory times, these encoded memories consist of intrinsically unstable physical qubits. However, encoded information in the system is maintained for arbitrary timescales via active QEC. Active quantum memories form the foundational building blocks of the QuNET system, allowing us to maintain encoded quantum information for long periods – even decades – indeed, times long enough for most important real-world applications.

The second major element to the QuNET system is **PORTABILITY**. The underlying qubit technology developed by Turing allows us to package these active quantum memories into systems that can be physically moved from place to place.

If we consider the six major qubit technologies that are the most developed and have been most successful in attracting significant private/public sector investment, we see that the *infrastructure technology* required will make physical portability difficult, if not impossible.

- **Superconducting qubits** require operational temperatures of approximately 30mK (0.03 degrees C above absolute zero) in order to maintain operational coherence. This temperature requires dilution refrigeration technology which cannot be moved.
- **Ion-Traps (IonQ, NQIT)** require an almost perfect vacuum environment and dilution refrigeration technology to enhance the creation of a perfect vacuum at dilution temperatures.
- **Optical technology**: Physical alignment of optical technology is extremely vulnerable to physical motion and therefore requires extensive motional insulation. Optical sources and detectors also utilize superconducting technology that requires 30mK dilution refrigeration technology.
- **Quantum Dots**: Similar to superconducting technology – operates at 30mK dilution refrigeration temperatures
- **Phosphorus in Silicon**: Similar to quantum dots and superconductors – requires 30mK dilution refrigeration temperatures.
- **Anyons (topologically ordered matter)**: Similar to quantum dots and superconductors – requires 30mK dilution refrigeration temperatures. Anyons have not been reliably experimentally demonstrated.

The Turing QuNET technology, by contrast, is based on qubits implemented as nitrogen atom defects within a diamond crystal. The diamond crystal itself provides what is known as a *spin vacuum* substrate [AGP11]. Essentially, the diamond crystal itself provides the same isolation properties that an actual physical vacuum does for ion-trap technology. The operational temperature of the Turing system is 4K. While still cold, 4K cryogenic technology is far simpler than the dilution refrigeration systems needed for a 30 mK thermal environment. 4K cryogenic technology is so advanced that we are able to effectively launch these sorts of cooling systems into space [GRW+07]. In 2003, the Spitzer space telescope was launched by NASA. On board is 360 liter liquid helium cryostat needed to cool instrumentation to approximately 1.5K to look at faint heat signatures from astronomical objects.

The Turing chip set is an array of optically coupled nitrogen-defect qubits embedded within a diamond lattice. The chip itself consists of an etched silicon base, with a ultra-thin diamond wafer "glued" on top. The diamond wafer is doped with individual nitrogen atoms separated from each other sufficiently that they don't directly interact. Individual qubits are coupled to each other using a layer of integrated silicon optics that sits above the diamond layer. The details of the hardware design and fabrication processes can be found in a separate white paper.

Optical pulses are sent between individual nitrogen-defect qubits to enact multi-qubit gates. These optical pulses can, in general, be weak coherent states that are easily produced. The system geometry is spaced out and optimized to allow the control structures for both the NV and optical layer to be fabricated to high accuracy.

Figure 2: The Turing diamond based chipset within a portable 4K cryostat system. High operating temperatures and no vacuum systems makes the system significantly easier to physically move than architectures employing vacuums (Ion-Traps) or Dilution technology (Superconductors).

Shown in Figure 2 is the device itself. On the right hand side is a rendering of the microscopic detail of each chip, with multiple qubit arrays (chip sets) connected to each other with fibre optic connections. Shown on the bottom left is a single chip set (which can contain approximately 50 physical qubits). On the top left is a commercially available 4K liquid helium cryostat similar to the device currently used in our prototyping laboratories. The cryostat itself is about the size of a coffee machine, with the cylindrical box containing the chip-sets themselves. The rest of the cryostat (the vertical part of the device) is a helium recycling system that allows us to re-condense the liquid helium as it evaporates from the sample chamber. The device consumes approximately 3 to 4 liters of liquid helium over the course of about 12 to 18 months. Losses are mainly though leakage in the closed loop recycling system.

The cryostat system is designed to accommodate multiple quantum chip sets. All the individual physical qubits in the system are connected to each other using integrated silicon optics and fibre connections, and multiple chips can be easily connected to allow quantum information to interact across multiple chip sets. As we will discuss in the next section, each chip set within the device will be used to create a single *encoded* piece of quantum information that will have a long enough coherence (memory) time to enable the operation of the Turing QuNET.

4.1 Long-lived quantum bits

Active quantum error corrected devices are the cornerstone of scalable quantum computation and communications technology. There has been significant discussion within the community related to so-called Noisy Intermediate-Scale Quantum (NISQ) applications [Pre18]; that is, applications small enough to not require computational or communications systems to be encoded using QEC. However, no NISQ application has yet been identified of scientific or commercial relevance that can be performed without employing QEC protocols. QKD is a case in point. One of the major implementation-related drawbacks in current QKD technology arises from the physical errors that occur in the transmission of quantum information and the limited ranges and rates available to optical, free space [UTSM+07, MHS+12], and satellite based quantum networks citeSD-Tang:2016aa,SD-Takenaka:2017aa,SD-Yin:2017aa due to these errors.

Active QEC protocols are central to the Turing QuNET architecture. By skipping any consideration of so-called NISQ applications, we develop a technological base that can be used to develop long-range, high bandwidth, and flexible QKD and quantum communications networks. In so doing, we lay the foundation for technology that can scale all the way up to massive, networked, and fully error-corrected quantum computing systems.

As mentioned above, the underlying technology of the Turing QuNET is a diamond-based qubit chip set. This chip set is designed to house an array of optically coupled qubits that are encoded using efficient error correction (QEC) techniques into a long-lived quantum memory. Turing has developed several advanced methods of QEC implementations on our chip sets. These methods are designed to optimize and reduce the physical resources needed to obtain a certain level of operational performance. The details of these methods are beyond the scope of this discussion. Instead, without loss of generality, we will use a particular, well known error correction technique to illustrate the basic operational principles of our chip sets and how they can be scaled up into the devices that will ultimately be used to construct the QuNET.

Surface code quantum error correction is a well known technique for active error correction that is being developed by several of the major hardware vendors in quantum computation [FMMC12]. The basic idea is that a square 2D array of interacting qubits is needed to encode a *logical* piece of quantum information to extend its natural lifetime beyond the individual *physical* lifetimes of the constituent physical qubits. The basic schematic is shown in Figure 3.

The physical array of qubits is arranged in a 2D nearest-neighbor connection geometry, wherein individual physical qubits can couple to immediate neighbors to the north, south, east, and west as shown in the top left image in Figure 3. In the Turing chip set, we use integrated silicon optics and optic fibre to connect together the physical qubits, allowing more flexibility in the physical layout.

The Turing chip set is fabricated as a 2D array of silicon micro-cavities in the diamond crystal, each of which holds an individual nitrogen atom. The top right image in Figure 3 shows a lithograph of the Turing chip set, showing a row of micro cavities. Each of the physical qubits in the chip set is separated by approximately 250μm horizontally, with each row separated by approximately 400μm. These are the physical dimensions for the first-generation Turing architecture, and were chosen so as to make control wiring and packaging simpler to engineer. We expect to be able to miniaturize the device by many orders of magnitude in subsequent generations. Shown in the bottom right image is the chip set itself and its physical size. The image on the bottom left in Figure 3 is a performance analysis of a surface-code QEC system that could be im-

Turing chipsets and benchmarked effective memory times.

Chip (physical qubit array)	Physical size (1st generation Turing chipset)	Quantum memory time
10 x 10 = 100 qubits	2.5 x 4 mm	22 milliseconds
20 x 20 = 400 qubits	5 x 8 mm	32 seconds
25 x 25 = 625 qubits	6.3 x 10 mm	18 minutes
30 x 30 = 900 qubits	7.5 x 12 mm	10 hours
35 x 35 = 1,225 qubits	8.7 x 14 mm	13.6 days
40 x 40 = 1,600 qubits	10 x 16 mm	1.2 years
45 x 45 = 2,025 qubits	11.3 x 18 mm	36 years

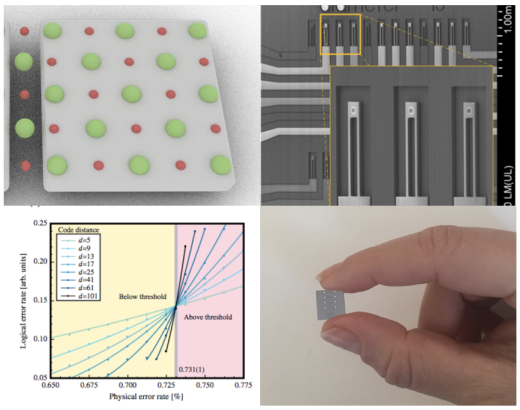

Figure 3: Expected resource overhead for a Quantum Memory Unit based on rotated planar codes [HFDM12]. A 2D nearest neighbour array of physical qubits can be used to encode a single *logical qubit* of quantum information. Provided the underlying hardware has a physical error rate associated with initialisation, measurement, single and two-qubit gates less than the fault-tolerant threshold of the surface code – approximately 0.7% – increasing the size of the lattice will exponentially increase the effective coherence time of the stored information. In the table we illustrate an example of a QMU built from a physical array of qubits that have a physical separation of 250μm, physical gate times of $t = 1\mu$s and physical error rates of $p = 0.1\%$. We specify the expected *logical* memory time at a *logical* fidelity of 99.99% for a hardware with stochastic, balanced qubit noise [DGSVM16].

plemented in the Turing chip set indicating how the natural lifetime of the system can be extended.

In the table illustrated in Figure 3 we detail the number of physical qubits, the physical size of the Turing chip set, and the amount of time we can maintain the coherence of a single piece of quantum information. Once the chip set contains approximately 900 physical qubits at a physical dimension of about 6.3×10 millimeters, we can extend the natural coherence time of a piece of encoded information and reliably store it for approximately 10 hours. The performance of the system scales exponentially. Consequently, if we roughly double the size of the chip set, the effective quantum memory time increases to 36 years!

It should be emphasized that these estimates are based on QEC techniques that are un-optimized for the Turing chip set. As noted earlier, Turing has developed QEC protocols specifically tailored for our hardware system. This substantially reduces the required physical qubit array sizes for a given memory duration. Performing error correction decoding on a system with biased noise [TBF18] can change resource overheads and it is possible for other QEC coding techniques to be invented that is compatible with hardware engineering constraints that scales better for the rotated planar code [Bom15, BVC+17, FGL18]. The key message here is that our diamond-based chip sets can be made into long-lived quantum memories with a physical size of order 1cm^2 that can be housed in an ultra-portable 4K helium cryo-stat. This is the system that will enable us to build a practical, cost-effective QuNET.

Illustrated in Figure 4 are the four primary technological properties that makes QuNET a game changer for QKD and quantum networking. The system developed at Turing allows for portability and long-lived quantum bits. But that is only of academic interest if the system cannot be manufactured economically.

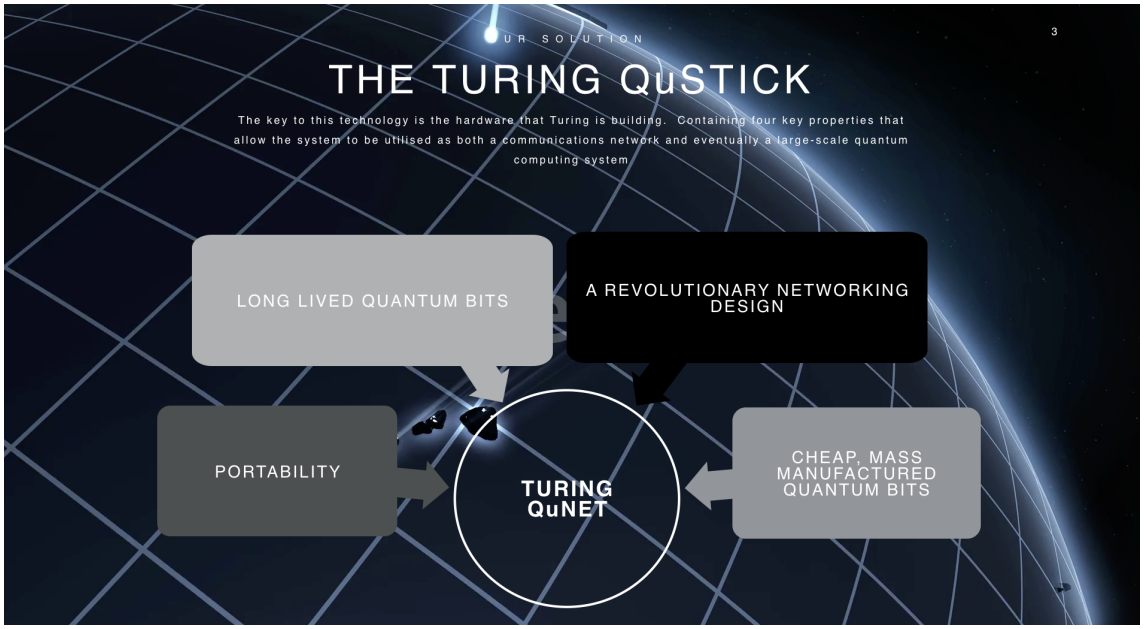

Figure 4: Four core elements that enable the Turing QuNET.

4.2 Cheap, mass-manufactured chip sets

Cost will always be a major issue for any quantum computing or communication/QKD system. While quantum technology opens up a plethora of possibilities in terms of computational power and communications flexibility, it cannot do so with simply a handful of qubits. Unlike the development of classical computation (where the competition for early computers was a room full of people with slide rules) or communications systems (where the state of the art was message exchange using carrier pigeon), quantum technology is competing with an extremely sophisticated and powerful classical infrastructure.

Many researchers have made the claim that building a quantum computing system or quantum communications network will be akin to other major scientific projects such as the Large Hadron Collider (LHC) at CERN or an array of LIGO gravitational wave detectors for astronomy. However, in our opinion, this is not quite true. While LIGO and the LHC are extremely expensive scientific projects, in the case of the LHC there is only **one** and with LIGO there are only currently **two** units in the field. Quantum computing and communications systems will be ubiquitous in the future. We already know better than to believe, as IBM chairman Thomas Watson reputedly did, "*I think there is a world market for about five computers*".

Consequently, being able to mass manufacture physical qubits cheaply will be of paramount concern for any hardware developer in this space. And when we say *cheaply*, we really mean it. Physical qubits in a quantum system are often analogized with physical transistors in a classical information processing system. As you can see in Figure 5 the cost of individual classical transistors is low. While it is certainly not expected that the price-per-qubit (PPQ) will reach these levels anytime soon, we need to keep Figure 5 in mind if we ultimately want to make quantum technology at the scale needed to take full advantage of the power of quantum computing and communications.

The Turing system is designed with cost in mind. Specifically, for the first generation Turing chip set, we are targeting a PPQ of $1. It should be noted that PPQ needs to take into account the infrastructure and control systems as well as the physical qubit itself. For example, in superconducting systems, the required dilution refrigeration system is expensive – around $500,000. Additionally, the sample chamber for the dilution refrigerator is quite small in comparison to the "footprint" of a superconducting qubit. A commercial dilution refrigeration system can accommodate perhaps 100 such physical qubits. Consequently, at a minimum, the PPQ in this system would be $5000.

Figure 5: Exponential decrease of the cost of classical transistors over the past 50 years. Plot was constructed using data from [BDR00].

Shown in Figure 6 is a breakdown of the Turing chip set architecture and how far we need to reduce costs in order to hit a PPQ of about $1. Most of the infrastructure and control cost for fab-

CHEAP QUBITS

CURRENT COSTS

Each chipset has been costed out using currently available bulk purchase prices and can be compared to what we need to hit our $1-$10 per qubit target.

COMPONENT	PER CHIP (BULK)	TARGETS
SILICON CHIP	$250 USD	$25 USD
SILICON CHIP FABRICATION	$1,000 USD	$100 USD
SILICON COATING	$100 USD	$10 USD
DIAMOND CHIP	$4,000 USD	$400 USD
NITROGEN IMPLANTATION	$1,000 USD	$100 USD
DIAMOND COATING	$150 USD	$15 USD
PACKAGING	$1,000 USD	$100 USD
OPTIC FIBRE ARRAY	$20,000 USD	$2000 USD
OPTICAL DETECTORS	$5,000 USD	$500 USD
CRYOSTAT	$250 USD	$25 USD
OPTICAL SWITCHING	$2,000 USD	$200 USD
COMPUTER CONTROL	$10,000 USD	$1,000 USD
LASERS	$500 USD	$50 USD
TOTALS	**$45,250/chip**	**$4,525/chip**

Figure 6: Current and projected costs of NV chipset.

rication is on a per-chip basis. The number of actual nitrogen defect qubits within a single diamond wafer (and hence per chip) can vary without significantly altering the per-chip cost. Our current costings, which are based on custom technology for each component, is approximately $45,000 per chip for a system with about 50 physical qubits. Our targets are a global 90% reduction in costs for each component of the first-gen chip set once mass manufacturing and vertical integration are implemented. This will reduce the cost of a single chip set to approximately $4,500. The capacity target for a single chip set in this generation is approximately 1000 physical qubits, leading to an

active quantum memory that lasts approximately 24 hours [Figure 3], and averaging out to a PPQ cost of $4.50.

We expect costs to drop further, thanks to economies of scale and efficiency in the manufacturing processes, as well as by increasing the number of physical qubits per chip set. However, our initial pricing targets for all elements of the architecture should take us into the ballpark of a $1 PPQ. This projected PPQ allows us to scale the technology far more ambitiously than other systems.

5 Operating Principles of the Turing QuNET

The basic operational primitive of the QuNET system is an old technique for the long range distribution of information, informally known as a "sneakernet". While essentially everyone today is familiar with the principle of classical Sneakernets, many people are not familiar with the name https://en.wikipedia.org/wiki/Sneakernet. As encapsulated in this wonderful quote from Warren Jackson, the former director of the University of Toronto Computing Services (UTCS):

Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.

In a classical sneakernet, rather than sending information via radio waves or fibre optic cables, it's loaded onto storage memory and physically transported from source to destination. The viability of sneakernets in the classical world depends heavily on three practical considerations:

- Availability of cheap, high capacity classical memories
- Limited bandwidth and high costs associated with radio or optical fibre communications channels
- The need or lack of need for low latency communications

The first practical consideration for sneakernet communications has been solved. Classical memories have fallen in price and increased in capacity to an even greater degree than transistor price and density. In 2018, Toshiba released a single 3.5-inch hard disk drive with a capacity of 14 TB available for $550.

Let's contrast this capacity with fibre optic connections. In 2016 the new FASTER fibre optic communications link was brought online between northern California and Japan. Costing $300 million to deploy https://www.nec.com/en/press/201408/global_20140811_01.html, it has a total data capacity of 7.5 TB/sec across the Pacific. Consequently, it takes a full two seconds, utilizing the full design capacity of a $300 million dollar piece of telecommunications infrastructure, to transmit the data contained on a $550 device that could fit into a handbag. This $550 hard drive can be shipped via FedEx from northern California to Tokyo for about an extra $150 and arrive about twenty hours later. Using that approach, we could achieve the same bandwidth as the FASTER network utilizing approximately 36,000 hard drives shipped continuously back and forth. The total cost of this would be approximately $20 Million dollars for the hard-drives and $5 million per day in commercial shipping costs (which could be made significantly cheaper by using more dedicated cargo transportation systems).

The capital cost alone associated with the FASTER network is *fifteen* times more expensive than the capital cost of using hard drive-based sneakernet. The operational costs of the hard drive system would be comparable if not lower than these fiber-optic links. So why do we spend so much time and money on capital and maintenance-intensive classical communications infrastructure?

The answer is the third point noted above: **information latency**. When we transmit information from one side of the planet to another, we don't want to wait twenty hours for the data to arrive. A classical communications system that has that amount of information latency is unusable for most practical applications. Consequently, our classical infrastructure for communications is designed to operate with high bandwidth (capacity) at as close to the speed of light as physically possible.

Once we move into the quantum regime, we can get the best of both worlds – the infrastructure benefits of a sneakernet-based communications link without the latency problem traditionally associated with that approach. How is this done? The key to the Turing QuNET is **portable, long-lived quantum hard drives.** We call these quantum hard drives **QuSTICKs**, and they constitute the basic units of the Turing QuNET [DGSVM16].

A QuSTICK is a single, portable 4K cryostat that contains a large number of the active quantum memory chip sets described in the previous section. As we have noted, a chip set array of diamond-based qubits can maintain and store quantum information for time periods ranging from a day to many decades, depending on the number of physical qubits in the chip set. In Figure 3, a $1cm^2$ chip-set has enough QEC hardware to protect the information for approximately 24 hours. A QuSTICK consists of hundreds, potentially thousands, of these chip sets in a common cryogenic environment, packaged into a single device. A QuSTICK is the quantum equivalent of a portable memory stick. However, a classical memory stick and a QuSTICK transport very different things. A memory stick carries classical data; a QuSTICK carries quantum entanglement. This distinction is the key to the power of the Turing QuNET.

Shown in Figure 7 is the basic structure of the QuSTICK communications link. Two parties, Alice and Bob each have a QuSTICK. The device itself contains multiple Quantum Memory Units (QMU's). Our hardware design calls for approximately 100 QMU's in the first-generation QuSICKS. Each QMU is a NV-diamond chip-set containing sufficient physical qubits to create a long-lived quantum memory of some defined timescale. Referencing back to Figure 3, if the desired memory time is about a day, each QMU chipset would consist of approximately 1000 NV-qubits. If we wanted each QMU to protect its respective quantum information for a year, each chip-set would contain approximately 1,600 NV-qubits. In the figure, and below in this document, we refer to the productized version of a QuSTICK as a QuBE. A QuBE includes all necessary packaging, connectors, monitoring sensors and readouts, etc.

5.1 Transporting Quantum Entanglement.

Entanglement is a unique property of quantum mechanics that has no classical analogue. Once referred to by Einstein as "spooky action at a distance", entanglement is the ability for quantum particles to remain linked after they have been interacted together regardless of physical separation or physical obstruction. Unless quantum decoherence occurs (the tendency of quantum particles to lose their "quantumness" over time due to interactions with the larger environment or deliberate measurements of their state), there is no evidence that quantum entanglement can be disturbed, blocked or otherwise tampered with. According to the basic principles of quantum mechanics, if two quantum particles are entangled and isolated well enough from the outside world, they can be transported to the opposite sides of the observable universe, with all the stars, planets and black-holes between them, and the entangled connection will remain undisturbed. It is this entangled state that will be transported by the Turing QuNET. Entanglement isn't information, but rather a quantum resource that can criss-cross the globe and be used as a consumable for quantum related protocols such as QKD.

In Figure 3 we illustrated a simple two-party point-to-point connection that is possible with QuSTICKs. Turing fabricates two QuSTICK units that house some number of quantum memory chip sets (recall that the first-generation Turing design calls for a minimum of 100 QMU's per QuSTICK). Each QMU is designed to maintain its integrity for some predetermined amount of time using active QEC protocols built into the QuSTICK. Both QuSTICK units start out in the same room. Because *every* physical NV-qubit within a QMU (and between QMU's) are optically connected to each other, it is no more difficult to interact NV-qubits in two separate QuSTICKS than it is to interact NV-qubits that exist in the same QMU. This allows us to create entangled states between QMU's in separate QuSTICKs by simply connecting them together with a suitable fibre optic connection.

The most basic entanglement protocol is to match up each individual QMU in QuSTICK-1 with a partner QMU in QuSTICK-2 and sequentially create an entangled state between each pair of

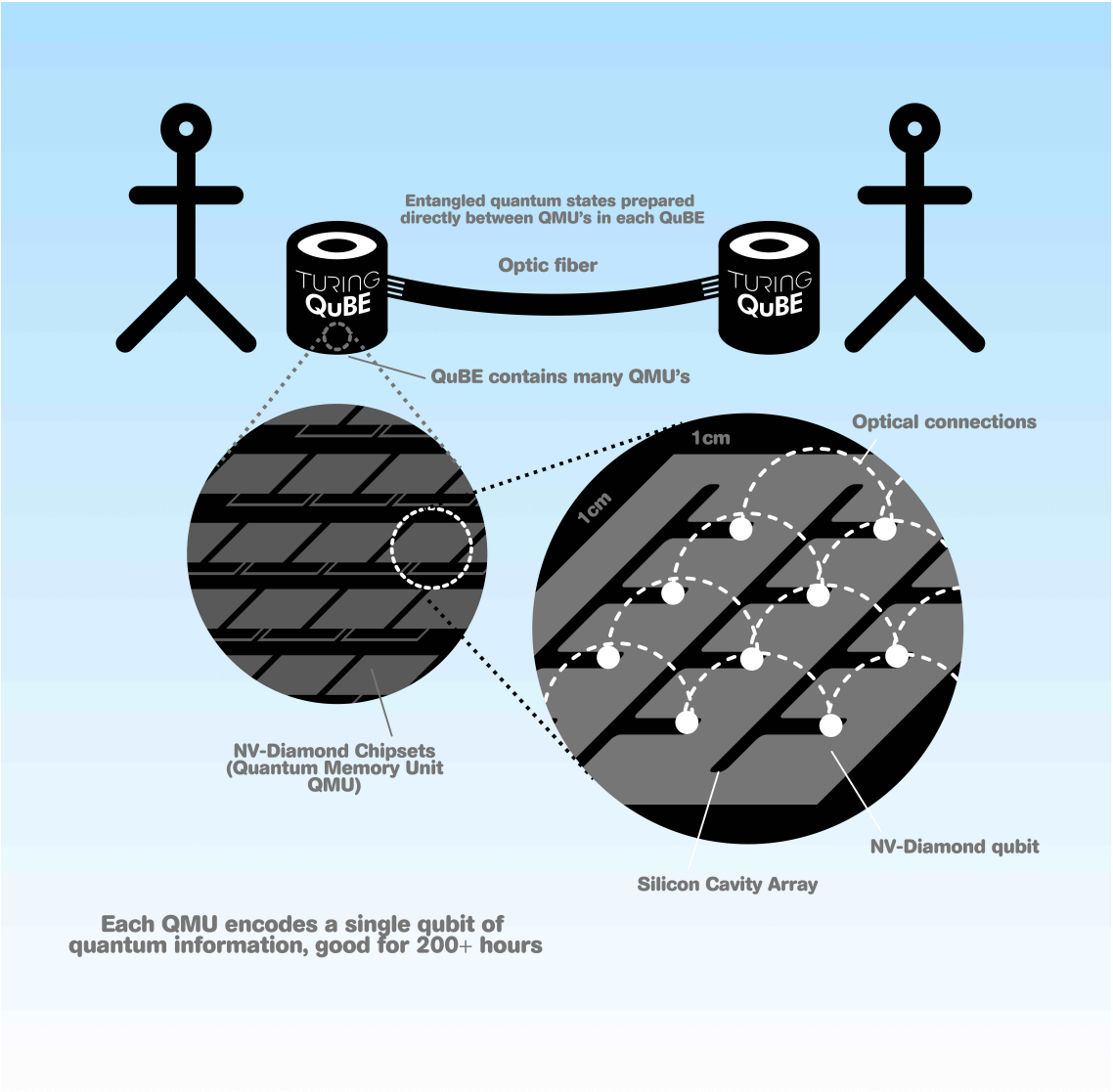

Figure 7: A QuSTICK is a portable device containing a collection of active QMU's. The QuSTICK is to be portable and ideally with an optical interface to the physical qubits comprising each QMU. In the case of colour centers, which in general have optical transitions available, interacting physical qubits within QMU's and between QMU's separated by a small amount of optic fibre should be nearly identical. Hence two separate QuSTICK units can be coupled together, entangling respective QMU's in each unit into Bell states.

QMUs. Once this is done, the optic fibre physically connecting the two QuSTICKs can be removed and the internal error correction will preserve the quantum entanglement up to the time specified by the physical number of NV-qubits inside each QMU.

Each QuSTICK unit has a finite number of QMU's, but there is no limit to the number of QuSTICKs that can be manufactured and deployed in the QuNET. For each pair of QuSTICKs we execute this pairwise entangling protocol and then then load half of them into a transport vehicle. Shown in Figure 8 is a scenario where Turing's own manufacturing factory is "home base" and we are preparing long-lived entangled states with a set of QuSTICKs that remain at "home" and partner sets that are physically moved to a different location.

It is important to note that when the entanglement is initially prepared, we do not need to know the destination of the transported QuSTICKs or their eventual application. The application could be highly classified QKD distribution or it could be a very public scientific experiment. The

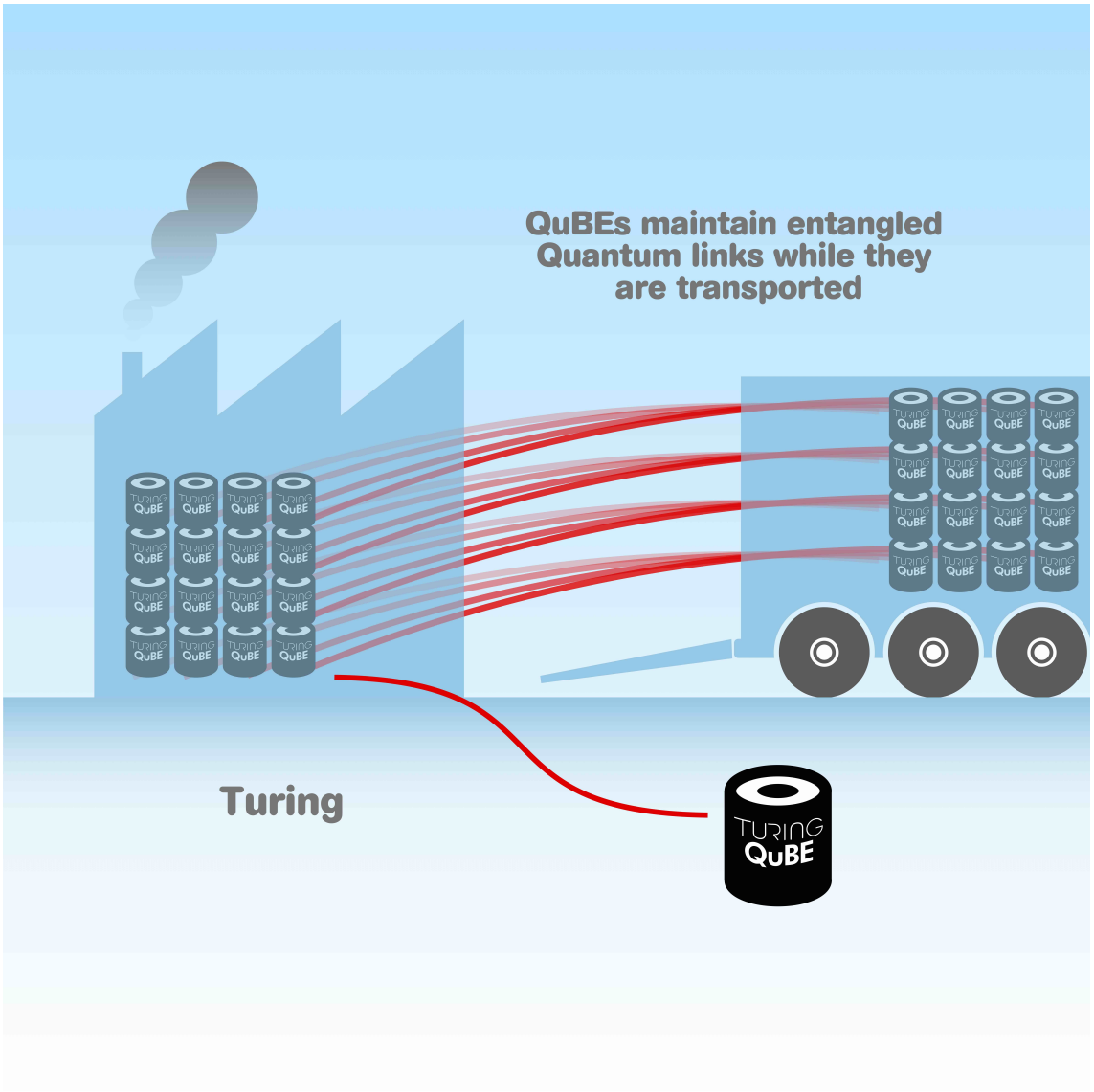

Figure 8: QuSTICK units can be manufactured and entangled locally. Due to the portability of the units, they can be physically shipped. This distributes the entanglement directly, without transmitting quantum information via free space or optic fibre over long distances. The internal error correction of each QMU will maintain the coherence of the entangled states during the physical transport.

only thing we care about is that the QMU's inside each QuSTICK have sufficient quantum memory time to get to their destination with their entangled state intact.

Once the entanglement is prepared, the entanglement links are *automatically* preserved while the devices are being physically moved. The internal error correction protocols for each QMU ensures that errors produced by the physical NV-qubits themselves or the act of physically moving the system are effectively corrected. The portability of the Turing NV-architecture is crucial to this. Other possible quantum computing systems are generally not portable. This is due to the heavy infrastructure requirements associated with extreme cooling or the maintenance of perfect vacuum states within the device. As mentioned earlier, the 4K operating environment of the Turing architecture is easily portable, and the diamond crystal itself acts as what is known as a spin-vacuum substrate for the NV-qubits. A spin-vacuum is an environment that looks like a complete vacuum to an individual qubit, even though the qubit is sitting, tightly packed, inside a physical crystal lattice.

Not all QuSTICKS have to go to the same destination. Once the entanglement is prepared at home base, a subset of QuSTICKS may be transported to destination A, another subset to destination B and so on. In fact, destination A may be 5km up the road, while destination B is half way around the world. Destination C might use low-cost cargo shipping to receive their QuSTICKs, while destination D's need might be for fewer units sooner, and so shipment by air is preferred (as shown in Figure 9). The flexibility enabled by portable quantum memories allows for dynamic allocation of both QuSTICK resources and the use of whatever physical transport method is appropriate to support the final application. This is in contrast to the "one size fits all" approach used by infrastructure-intensive communications systems like satellites and repeaters. Such systems operate in exactly the same way regardless of the demands of the ultimate application.

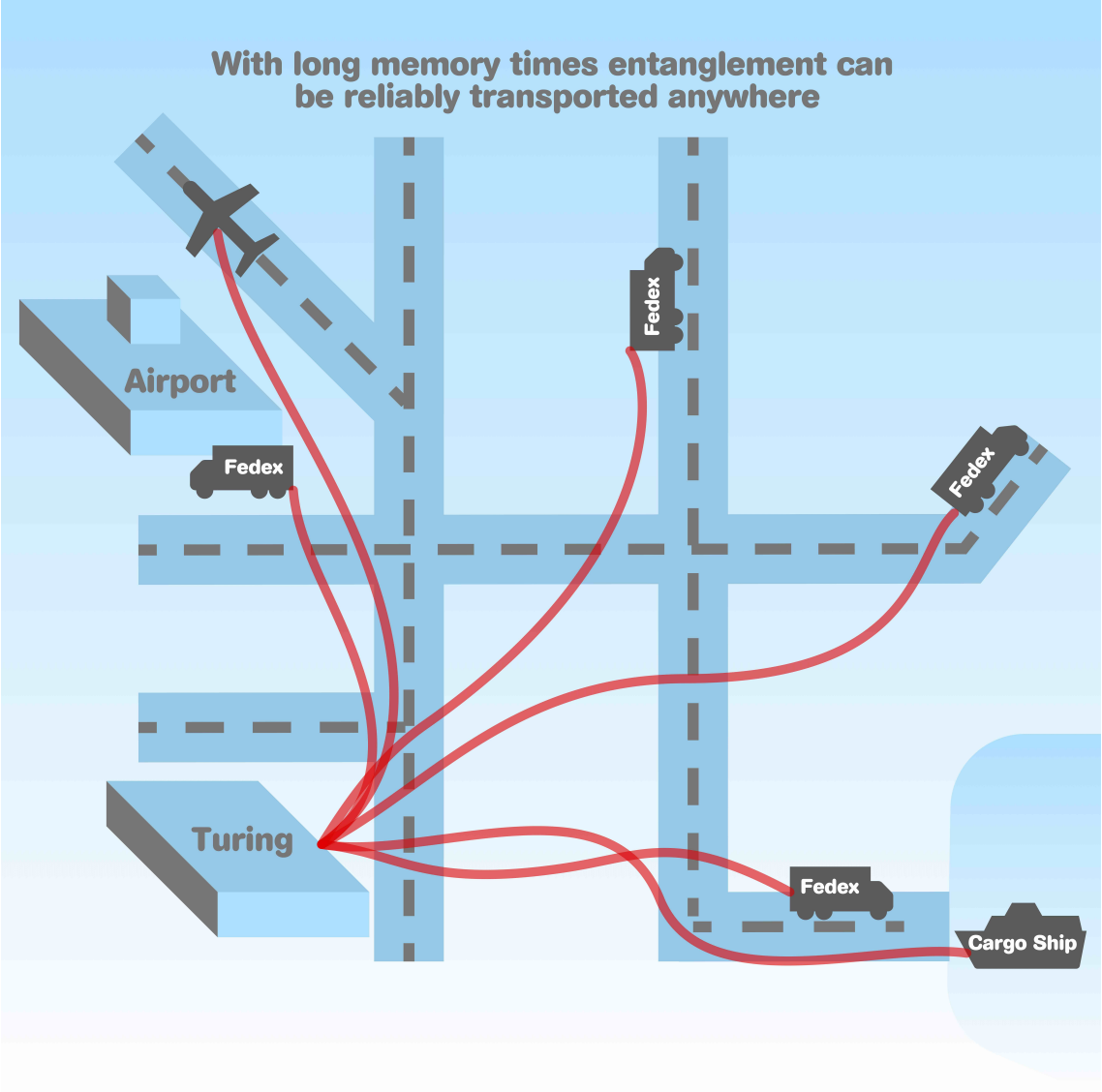

Figure 9: Entangled QMU units can be distributed anywhere that conventional shipping transport can go. A distributed, high-fidelity quantum network over global distances can be established by simply putting subsets of QuSTICK units onto different transports and moving them to separate locations.

5.2 Using and re-charging QuSTICK quantum entanglement

We have described a system where a finite amount of entangled quantum states is maintained by a set of QuSTICKs for a predefined period of time. This period of time needs to be sufficient to entangle the QuSTICKs directly together at the source and physically transport them to the

destination. This mechanism now mimics the behavior of a quantum satellite or quantum repeater system, but with a crucial difference that ultimately makes the QuNET system far superior from a practical standpoint:

- The sender and receiver now share a persistent, high fidelity entanglement link that **does not have to be used immediately**. Quantum satellites and ballistic repeater designs generate entanglement using photons. They are not designed to "store" the entanglement for future use – the entanglement needs to be consumed for some type of communications protocol immediately upon receipt.
- The total amount of entanglement a destination shares with home base is dependent only on the number of QuSTICKs used. If only 100 entangled qubits are needed for an application, a single QuSTICK will suffice. One million entangled qubits would require 10,000 QuSTICKS.
- The entanglement QuNET provides is **by design** ultra-high-fidelity. All of the numbers in Figure 3 assume that each QMU can maintain an encoded piece of quantum information for a given period of time **to a fidelity of 99.99%**. Increasing the fidelity even further requires only a small increase in the size of the QMU. If a target application requires a fidelity of 99.999999999% (say for direct connections between fully fault-tolerant quantum computing systems), simply use a few more physical qubits per QMU.
- QuNET's entanglement connections can reach global distances (Alice's QuSTICK in Sydney can be connected to Bob's QuSTICK in London using conventional transportation services).
- The persistent entanglement I share with you can continue to move before it is used. Alice may not want to use this entanglement in Sydney, but instead transport it to the desert in central Australia. The entanglement link with London will continue to move with Bob as long as his QuSTICK provides sufficient error-correction for the additional time required.

5.3 Using SWAP entanglement

Another extremely beneficial aspect to Turing's QuNET design is the ability to **SWAP** entanglement [ŻZHE93]. What does this mean?

In Figure 9 we assume that all QuSTICKs that will be physically transported are initially physically connected to partners sitting at their home base. This creates a star-network structure, where the home base acts as an endpoint node for all of the entanglement links that are physically distributed outwards. But suppose two distributed QuSTICK units want to share entanglement directly? Do they need to physically meet and have their optical links hooked up? No, instead they can ask for a entanglement SWAP protocol to occur.

Let us take three parties: Alice, Bob and Alan (Turing). Alan is at Turing HQ in Berlin, and, as in the previous example, Alice is in Sydney and Bob is in London. In total there are four QuSTICKs. Alan has initially manufactured all four, and we will call them QuALICE, QuBOB and QuALAN-(1,2). Before putting QuALICE and QuBOB onto airplanes or ships, Alan entangles all of QuALAN-1 QMU's to the QMU's in QuALICE, and similarly between QuALAN-2 and QuBOB. QuALICE and QuBOB are then shipped off to Sydney and London respectively. Again, at this point, nobody has decided exactly how they are going to use this entanglement; but (depending on the amount of memory time provided by each QMU in each QuSTICK) they have a fixed time during which they can use it for anything.

Once QuALICE and QuBOB arrive, Alice and Bob only share entanglement with Alan directly. However, at some point, Alice and Bob decide that they want a direct entanglement connection with each other (for example, if they are intending to establish some cryptographic keys).

Instead of physically moving again, Alice and Bob can request an entanglement SWAP from Alan. Alan will then connect QuALAN-1 and QuALAN-2 directly together (since he still has physical possession of both units in Berlin) and entangle each of their QMU's. If Alan then measures each of the QMU's in his two QuSTICKs in a particular manner ("measure" just means reading out the quantum state of each QMU sequentially), the entanglement he initially shared with Alice and

Bob is SWAPPED to them. By performing this operation, Alice and Bob can share **directly** two QuSTICKs of entangled QMU's without ever having to physically meet each other. They can then proceed to perform whatever communications protocol they wish. This SWAPPING protocol is illustrated in Figure 10

The above example highlights an important point. The physical distribution of a set of QuSTICKs creates an *initial network topology* for their entanglement. However, this topology can be modified after the fact to create direct entanglement connections between parties that were never initially in the same distribution channel. The physical distribution network can be thought of as a network graph. Initially, each node is the physical location of a QuSTICK, and each edge connects the node where a QuSTICK began its journey to a node where it ended its journey. Entanglement SWAPPING allows us to change the structure of the graph (i.e. change where edges are and are not) without having to move the QuSTICKs again.

This kind of flexibility is simply not possible in classical networking design. It would be as if a classical data packet could be sent **directly** between the Australia and Iceland despite there being no direct telecommunications link between those two countries. Classically, the data packet would have to first travel from Australia to Singapore, then perhaps from Singapore through routers in west Asia, the Middle East, central and northern Europe, the U.K. and finally to Iceland. At any point in its journey the packet might be intercepted, copied, lost, tampered with, inspected, corrupted, or accidentally routed to a completely inaccurate destination.

With entanglement SWAPPING using the QuSTICK network, we can define a network topology first, before it is used, and then later define direct entanglement connections between parties that were never in direct physical contact.

This opens up a whole new world in network design theory – and it is only possible in the QuNET system.

5.4 Entanglement depletion and network persistence

The use of the entanglement that is prepared between two parties to actually perform a protocol is discussed in more detail in a later section. But in all cases the following set of steps occur:

- The parties to the protocol first configure the entanglement network, through SWAPPING, to match the requirements of the protocol.
- Each party measures the *logical state* of each QMU by physically measuring every physical qubit that QMU comprises. These measurements can occur in many different ways (technically referred to as the *basis* in which the logic state is measured). Measuring a QMU results in a yes/no answer (and the QMU only ever gives yes/no answers). Choosing a different *basis* to measure the QMU is akin to asking a different yes/no question.
- The parties in the protocol announce, publicly, over a classical communication system, what *QUESTION* the QMU's are being asked. i.e. each party only announces the BASIS they chose to measure their QMU's in. They never reveal the ANSWER the QMU's gave them.
- All the classical *ANSWERS* between the QMU's are now completely correlated due to the initial quantum entanglement the parties shared.
- The entanglement carried by each QMU has now been destroyed or *consumed* to perform the protocol and is no longer present to use again (although, of course, the QuSTICK itself isn't affected and can be re-entangled any number of times).

What happens when the collection of QMU's within a QuSTICK has been depleted? The entanglement resource of a QuSTICK is analogous to a battery that carries a fixed amount of charge. A battery is a physical unit that can be used in many different ways and moved around to different devices; but it contains a resource that is finite. Once that finite resource has been exhausted, it must be replenished again from an appropriate source. In the case of a battery, it's recharged from via electrical outlet. In the case of a QuSTICK, the replenishment comes from another QuSTICK.

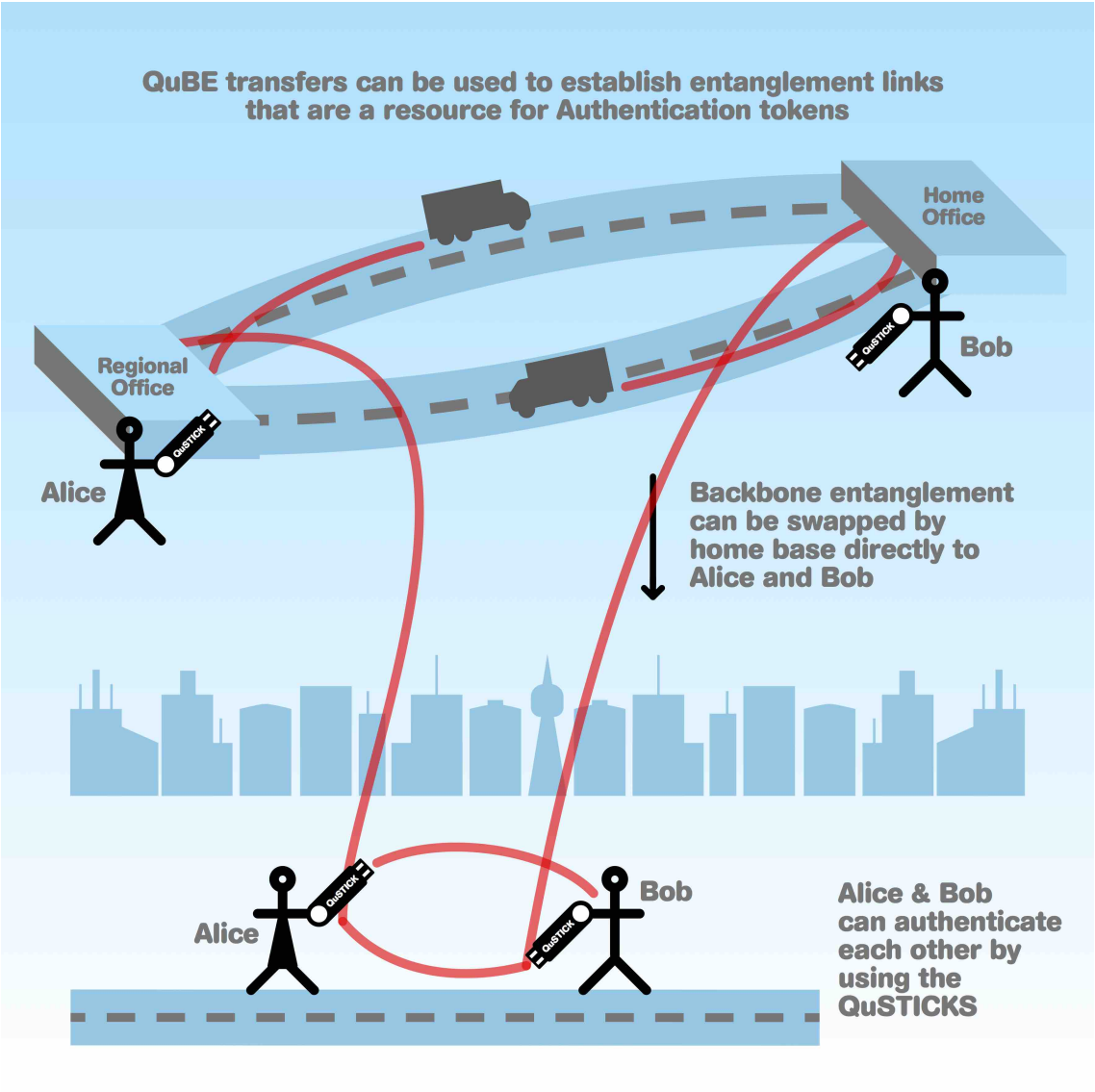

Figure 10: A backbone network of physically shipping a large volume of QuSTICKS between a home office and regional office can generate continuous Bell states between these two locations. Alice and Bob can take QuSTICK units from the two respective offices and travel to a third location where they meet. If an authentication protocol is needed – because Alice and Bob have never met and need to know they have come from the same organisation – the entanglement in each of their QuSTICKS can be SWAPPED by their respective offices such that they now share Bell states together. These Bell states can then be used to violate a Bell inequality, authenticating that their quantum resource was given to them by the same source.

Turing HQ may be the source of the physical QuSTICKs, but is not, in general, the source of entanglement. Returning to the battery analogy, the Tesla's gigafactory may be the source the physical batteries, but it isn't the source for the battery *energy* in your Tesla vehicle.

Network connections between QuSTICKs can be established whenever and wherever QuSTICKs can be physically connected to each other. And, as discussed above, via SWAPPING protocols, entanglement connections can be reconfigured over very long distances. And while the system is certainly designed to be portable over global distances, we can also exploit the "small-world phenomenon" [Mil67] to construct long range entanglement with only relatively short physical movement of QuSTICKS.

Once the entanglement in a QuSTICK is consumed and depleted, the owner or operator of the QuSTICK simply has to physically find another QuSTICK that has some entanglement "charge" left. If Alice has no entanglement left, she could go find Charlie (who hasn't consumed his entanglement yet) and physically connect her QuSTICK to his and re-entangle all the QMU's. At this point, Alice has reentered the entanglement network. Once she has, she can use SWAPPING to establish connections to another network member to perform more communications protocols. There is no need for her to reconnect to a central entanglement hub or a distant source that would require her to ship her QuSTICK to the other side of the planet and wait for it to return.

Entanglement is a consumable resource, and may ultimately end up a marketable commodity in a future *entanglement based economy* [DGSVM16]. But, unlike a battery, the entanglement "charge" is a free resource that comes as a consequence of two QuSTICKS being in the same room.

The processes described above enable a persistent network to exist. Even though users will consume QuSTICK entanglement whenever they use their device to perform a quantum communications protocol, that networked entanglement can be replenished easily. Again, the nature and structure of QuNET allows us considerable flexibility in network topology. We can essentially design a two-stage embedded network: One stage in which the structure of the network is determined by the physical interaction of the users who have access to physical QuSTICKs, and a second stage in which quantum information flow is dictated by the structure of the underlying network. This is inconceivable in the classical world of information networking. We can only begin to imagine what remarkable things are possible with this kind of technology infrastructure. We can be certain that it will unlock capabilities well beyond the basic ultra-high security quantum communications applications that we are discussing here.

6 QuNET protocols

Many different communications protocols can be enacted using QuSTICK entanglement and the QuNET. The principal determining factor for what can be done with this technology is the number of units that can be physically manufactured. As production increases, and the total number of QuSTICKs in the world grows, new communications applications become possible. One of the major benefits of Turing's design is that **the entire network is expandable**. We don't replace a QuSTICK every time we a new one is manufactured; we simply add it to the network to increase overall capacity. Generation one of the QuSTICK design will be compatible with future generations. New design iterations will primarily increase quantum data capacity (as is the case with classical memory chips), and a QuSTICK-based communications network will become more complex, higher bandwidth and more flexible as additional QMUs are deployed.

We will discuss each class of quantum communications protocol in turn, beginning with authentication – which requires the fewest QuSTICKs – and progressing through applications that require increasing numbers QuSTICKs, ultimately arriving the most complex application: a full fledged Quantum Internet [WEH18] capable of connecting together fully fault-tolerant quantum computing systems.

A central principle (violation of Bell's Inequality [CHSH69]) underlies each of these protocols, and the non-specialist reader may want to take a moment to review the fundamentals of quantum entanglement and its utility in secure communications. Many thorough treatments can be found online and elsewhere (e.g. the Wikipedia article on Bell's Theorem). Appendix A below also describes how a violation of Bell's Inequality can be used to confirm the existence of entangled states at two locations, along with the associated calculations.

6.1 Authentication protocols

An authentication protocol is generally a small data-exchange mechanism used to confirm that a user is who he says he or she is, and that the information that is to follow will not be coming from a third party or impostor. It is used extensively in classical communications for applications that range from the routine (when I go to www.google.com, am I really talking to Google?) to communications that determine the survival of the human race (did the launch order received by

an Ohio-class nuclear submarine really come from the White House?) In Turing's QuNET system, authentication tokens will be one of the first applications due to the comparatively small number of QMU's needed [CS01].

The key question that the authentication process must answer is this: How do we ensure that messages are coming from the requisite trusted source? Turing's solution is similar to physical authentication tokens that some bank customers are given to access their accounts for high-value transactions. These classical authentication tokens utilize a variety of methods to ensure that the server (e.g. a bank website) and the client (the person with a bank account) can share a secret message that can be used to authenticate a log-in session. Some utilize static password tokens, some use synchronous dynamic tokens and some use a type of "call and response" method. There are even services that use a smart phone APP to verify the phone's SIM card and then generate a one-time key. All of these techniques are vulnerable to security flaws that present themselves if: 1) the implementation is not done perfectly, 2) the token is lost, stolen, or duplicated, or 3) if some classical calculation that was assumed to be computationally difficult is compromised in some way.

Turing's solution uses the same basic principle as the classical authentication token – the client and server share a physical object – but, because our QuSTICK units are quantum in nature, we can use the laws of physics to ensure complete security of the authentication protocol. The underlying idea is that two parties, Alice and Bob, can violate a Bell inequality if and only if they share an entangled Bell state. That is, two QuSTICKs that are entangled will behave differently when queried than two QuSTICKs that aren't entangled.

To illustrate the protocol, we stipulate that our two parties share a finite number of Bell states between them, with one half of the Bell states contained within Alice's QuSTICK and the other half within Bob's QuSTICK. This shared Bell state will typically be prepared ahead of time in QMU's inside a QuSTICK system by a third party and then distributed to Alice and Bob. That is, Alice and Bob have received their respective halves of the entangled state from some other QuSTICK source of entanglement. The source could be a much larger backbone quantum network that is distributing long-range entanglement via the SWAP mechanism as described in the previous section. Alternatively, the physical QuSTICKS themselves could have been delivered to Alice and Bob by a third party or parties. However, as we will see below, this third party source of QuSTICKs or entanglement does not need to be trusted by either Alice or Bob.

If Alice and Bob share Bell states between each other, then they can perform what is known as a *Bell violation test*. A Bell violation is the quintessential test of non-local properties of quantum mechanics and has been verified experimentally many times since the early 1980's to test some of the most foundational principles of quantum theory. We will briefly summarize the basic test below. More detailed calculations can be found in Appendix A. In a nutshell: The Bell parameter, S is calculated by performing measurements over a shared Bell state. In classical theory, this parameter, S must be ≤ 2, while in quantum theory, $S = 2\sqrt{2}$ [CHSH69]. Hence to confirm that Alice and Bob indeed shared a quantum state, we must determine if $S > 2$.

In Appendix A we find, in a basic error analysis, for $M = 100$ shared Bell states between Alice and Bob, the standard error is approximately 0.4, implying that with 99% confidence, $S > 2$. Hence 100 QMU's in each of two entangled QuSTICKs can be used to violate a Bell inequality with very high confidence, guaranteeing (assuming that quantum mechanics is correctly describing the behavior of the QMUs) that Alice and Bob do actually share entangled states. Each QuSTICK is designed to distribute extremely high fidelity Bell states, and any possible hardware or implementation inaccuracy is addressed with the internal error correction of each QMU. As a result, the primary source of error in this example is sampling error.

6.1.1 Real world example

Let's consider how we might use QuSTICK technology to authenticate arguably the most important communications protocol we can imagine, namely a launch or abort code to a submarine-based nuclear missile.

Clearly, the order to launch or abort a launch needs to come from the relevant government or

military authority; and the actual physical duty to launch or abort is either the responsibility of the captain of the submarine or an on-board automated system attached directly to the missile. Being able to reliably guarantee that an order has come from the right place in this instance is a matter of life or death.

We can consider a QuSTICK-based authentication system that is physically integrated with the missile system (literally a QuSTICK containing $M = 100$ QMU's, or likely more for such a critical application). Prior to deployment of the submarine, each QuSTICK unit is entangled with a partner unit at central command. Each QMU is designed to have a sufficient level of quantum memory for the duration of its deployment. This can be years for highly error-corrected QMU's at a desired ultra-high fidelity. These QuSTICK units will then maintain the two-party entangled states until they are required for authentication.

Let's consider the scenario where a launch or abort order for a missile has been issued from the White House. At the point in time when the order is transmitted, all the White House QMU's are randomly measured according to the protocol described above. The information related to how each QMU was measured and the result of that measurement is transmitted *classically* as an initial authentication header to the actual message.

Transmitting the data unencrypted does not pose a security concern for the authentication protocol itself and can be intercepted or modified by an adversary, without creating a false-authentication event. An adversary could alter this data to *prevent* authentication, but they could not falsely identify themselves as the White House.

The QuSTICK QMU's that exist on the submarine do not need to be measured simultaneously with the QuSTICK QMU's at the White House. Instead, they are read once the classical authentication header is received by the submarine, and the Bell parameter S is then calculated. If the submariners find, with arbitrarily high confidence, that $S > 2$, then whoever generated the authentication header *MUST* have had access to the partner QuSTICK that was initially entangled with the submarine QuSTICK. Assuming that the partner QuSTICK, which should have been secured at the White House or the Pentagon for the entire time, has not been physically stolen by an adversary, the submariners can confidently conclude that the message has come from the appropriate, authenticated source. Physics dictates that there is no other circumstance that could result in a Bell parameter greater than two.

The only ways the Bell parameter, S, can be "faked" by an adversary are as follows, with their appropriate countermeasures:

- One of the QuSTICK units is physically stolen. This can be mitigated by always preparing the entanglement between the two units in a secure physical location and then *only* ever moving *one* QuSTICK unit from this secure home base. In the above example, if the submarine QuSTICK is physically modified, it simply cannot be used as an authentication device. In fact, there is no way to successfully compromise either of the two QuSTICKs. Either a device is untampered with or it is simply unusable; tampering cannot be used to send a fake authentication signal.

- An adversary can intercept the transmitted classical data stream and try to infer the quantum state of the submarine QuSTICK to transmit a false authentication signal (a "man-in-the-middle" attack). This would potentially succeed if the adversary somehow *knows* the initial shared Bell state $|\psi\rangle_{AB}$ used by Alice and Bob. A simple way to mitigate (if not eliminate) such a threat is to randomly select one of the four possible Bell states (see Appendix A) when the initial entanglement is prepared for each QMU and hard code these random choices within the QuSTICKS themselves. This would be done before a QuSTICK is deployed to the field and not shared after this initial entanglement. This provides a secret reference frame for each QuSTICK that would prohibit an adversary from falsifying an authentication signal if they perform a man-in-the-middle attack on the classical data stream that is needed for the authentication protocol.

- Denial of service. The adversary could simply block the classical signal exchange between the two QuSTICKS or could steal one of the QuSTICKS and deactivate all of the QMU's. There

is no mitigation to an attack of this kind, but it will never result in a false authentication signal being sent to the submarine.

- Failure of the QMU's inside the QuSTICK's themselves. For a variety of reasons, the encoded quantum state within a QMU may simply collapse. This could be caused by a hardware failure or a sufficiently large external perturbation of the system due to a variety of possible factors. Like the previous instance, this simply makes the QuSTICK itself inoperable. This will not cause a false authentication event because entanglement is no longer being shared and consequently cannot produce a Bell violation. However, a legitimate attempt to authenticate will fail in this situation.

There are other possible ways to "tamper" with the exchange of the authentication exchange between the two QuSTICKS, but these can be reliably mitigated using standard privacy amplification techniques and protocols seen within the Quantum Key Distribution literature. However, as noted before, the fact that we are using ultra-high-fidelity, error-corrected quantum memory units means that many of the classical error-correction protocols used in other quantum key distribution schemes are not needed. This allows us to perform authentication protocols with fewer QMU's and less overall physical resources.

Being able to "spoof" an authentication token, for example from the white house, is only possible if the spoofer literally possessed both the partner QuSTICK to the submarine and control of the classical side channel that is used to reconcile the protocol. Given that QuSTICKS are physically entangled at home base and that the spoofer would need to gain possession of the QuSTICKS that would ostensibly never leave home base, there would need to be an extraordinary event such that this physical device would be compromised. The simultaneous need to take control of whatever classical channel is used between the submarine and home base adds a second layer that needs to be compromised before a fake authentication signal could ever be sent. Additionally, QuSTICKS is compatible with more complex secret sharing protocols. Quantum secret sharing protocols are multi-party generalisations of the authentication protocol, where rather than two-units, three or more units can be used. In this event, all units have to be used in conjunction to produce a secure authentication token. This could, for example be the submarine, the white house and the pentagon. For an authentication session to be valid, all three parties have to cooperate and reconcile measurements in unison. Consequently, if an adversary is looking to spoof an authentication token, they would now need to gain control of both home base units and the classical side channel, rather than just the one. This would make an already low probability event even less probable. If this level of security is still not good enough, the distribution can occur over more and more parties at home, requiring adversaries to take control or steal more and more QuSTICKS to compromise the authentication session.

The portability of QuSTICK units is what drives many of the potential applications of Turing's technology. While 100 error-corrected QMU's is a large number compared to what is currently available, Turing's system is designed to scale to this level almost immediately. The long-lived nature of our QMU's and QuSTICKs – and the ability to move them anywhere – allows for deployment of these units across a wide range of platforms and environments.

Whenever ultra-high security is needed, particularly for message authentication, the Turing QuSTICK is particularly valuable. The fact that quantum entanglement is prepared locally, at home base, before any of these units are physically deployed in the field adds tremendous security benefits. Such benefits are not available with quantum protocols based on the transmission of the entangled pairs. Why? To put this a different way: It is very difficult to detect if an eavesdropper is stealing a photon from an optical fibre or a photon flying through the air. It's a lot easier to know if someone has hit you over the head and stolen your briefcase that contains a Turing QuSTICK!

In the submarine example above, we exploit an asymmetry – the fact authentication only needs to happen in one direction – to modify the protocol so that the classical communication between Alice and Bob does not have to occur simultaneously. It turns out that this asymmetry exists in most practical situations. Another example is online banking: The banks' online servers have to confirm that a login is indeed coming from the correct person or device before access is granted.

In the case of the old Hollywood depiction of spies meeting each other in a cafe with a predetermined secret phrase referencing the weather in Scotland, authentication has to happen in real time, with their classical message exchange happening simultaneously, to protect against deceptive users. With most quantum-related security protocols, there is a minute fraction of a second that can potentially be exploited to compromise the system in the example given above. But ultimately this is greatly preferable to the current vulnerabilities of classical authentication or cryptographic protocols where much of the security is dependent on the "honest" behavior of loyal, well-vetted personnel.

As quantum entanglement is not obstructed by any known physical process, the links provided by the QuSTICKS cannot be disrupted or destroyed unless someone has physical control of the QuSTICK units themselves. Provided an unencrypted communications channel exists that allows communication with Alice and Bob, QuSTICKS can always be used to perform secure authentication.

QuSTICK-based message authentication will likely be the initial application of Turing's platform as it can occur with a small number of physical qubits. However, our system is adaptable. Initially, while QuSTICKS are sparse, we can run the QuNET system as simply a network for secure authentication. As we fabricate and deploy more QMU's and QuSTICKS, the authentication network grows until we hit a critical volume of devices, at which point they can be re-tasked to more complex quantum communications protocols, and the growth cycle starts again. Ultimately, as QuSTICK volume increases, we can run authentication protocols, QKD protocols, distributed quantum computing and communications and anything in between over a shared network that does not require segregated sub-networks for highly secure applications.

6.2 Key exchange

The next step after authentication protocols is key distribution using standard QKD protocols and additional QuSTICK units.

The QKD literature is now very rich, so we won't go into as much detail in this document on how QKD can be implemented. The basic process involves distributing a shared random bit-string that can be used to encrypt messages using strong, symmetric classical cryptographic protocols. In the most current, declassified standards, the U.S. National Institutes of Standards and Technology (NIST) states http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf:

The design and strength of all key lengths of the AES algorithm (i.e., 128, 192 and 256) are sufficient to protect classified information up to the SECRET level. TOP SECRET information will require use of either the 192 or 256 key lengths. The implementation of AES in products intended to protect national security systems and/or information must be reviewed and certified by NSA prior to their acquisition and use.

Hence, key exchange for strongly secure encryption requires key lengths that are roughly a factor of three higher than what would be needed for the authentication protocols described above. Note that this factor of three does not account for error correction or privacy amplification protocols that we will discuss below. Note also that the information exchange as a whole is not provably *quantum* secure. That is, AES encryption has not been proven to be susceptible to attacks from Quantum Computers.

Quantum Key Distribution protocols based on quantum entanglement (for example, the Ekert-91 protocol), in a broad sense, consist of the following steps:

- Two parties, Alice and Bob, first share a large number of Bell states
- Alice and Bob, using a process similar to the authentication protocols described above, randomly choose to measure their qubits by first applying a single qubit rotation to their half of the entangled state prior to measurement in the $\{|0\rangle, |1\rangle\}$ basis (see Appendix A). In the case of QKD, the single qubit rotations are randomly chosen from two gates, the Identity (do nothing) or the Hadamard gate.

- Alice and Bob then "reconcile" their results. Reconciliation requires Alice and Bob to classically communicate (on unencrypted classical channels) what single qubit gates they applied to their Bell states. Alice and Bob **do not tell each other the measurement value they obtained**, they just tell each other how they performed the measurement.
- On average, Alice and Bob randomly chose to measure their qubits *differently* 50% of the time. These 50% of Bell states that were measured differently are discarded.
- The remaining 50% of measurements (which are now classical bits) are correlated between Alice and Bob (that is, Alice and Bob both know the actual results of each other's measurement, even though they have not directly communicated this information). Hence Alice and Bob now have a shared set of classical bits that can be used as a key.
- From this set of shared classical information, Alice and Bob use a subset of the key's bit string to perform error-correction and *privacy amplification*. Error correction (which consumes part of the shared bit string) is used to help correct for imperfections in how the initial Bell states were shared. Privacy amplification involves Alice and Bob comparing a random subset of the shared key (they disclose publicly some of the shared bit string) in order to detect any possible interception or manipulation by an eavesdropper.
- By sacrificing a certain amount of the shared secret bit string, Alice and Bob can bound the amount of classical information available to any potential eavesdropper. Once that bound reaches a certain threshold, Alice and Bob can conclude that the remaining bit string that is shared is secure.
- The remaining secure bit string can be used to classically encrypt the message using AES or some other strong encryption protocol.

There has been a significant amount of work done to quantify how much "raw" key data needs to be generated in order to "distill" a appropriately secure key of some desired length that can then be used to encrypt the information. The ratio between the raw key and the "usable" key depends on the error rate in the system. In the security analysis for QKD, it is generally assumed that **ALL** errors are caused by eavesdropping rather than hardware imperfections. Shown in Fig. 11 is an example of such an analysis from (http://dx.doi.org/10.5755/j01.eee.21.6.13768). Please note that only qualitative trends from Fig. 11 should be inferred from this plot as, depending on hardware assumptions, the relationship between the "raw" key and the "secret" (or usable) key can change.

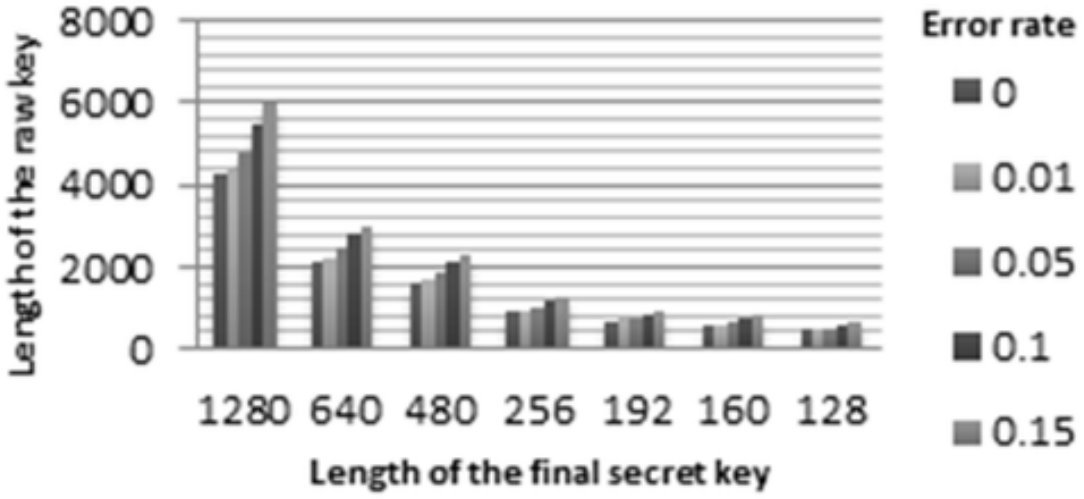

Figure 11: Relationship between the length of the "Raw key", the number of actual quantum states shared between Alice and Bob and the length of the "Secret Key", the number of quantum states shared that is actually used in the final key. As error rates increase, more and more of the Raw key is used for Error Correction and Privacy Amplification protocols.

In general, only 25% of the raw key material is finally used for the classical encryption protocols. This combined with the original 50% of the Bell states that need to be discarded in the QKD protocol implies that for a 256-bit classical key (appropriate for strong AES encryption), approximately 2000 Bell states need to be initially shared between Alice and Bob.

It should be stressed that, in the conventional analysis, a significant amount of the "raw" key material is sacrificed for error correction and privacy amplification in order to compensate for hardware errors that occur naturally (even though the protocol assumes that ALL errors are induced by eavesdropping). In the case of Turing's QuSTICKS, naturally occurring hardware errors are corrected – the QMU's in each QuSTICK are, by design, able to maintain ultra-high-fidelity Bell state entanglement. As a result, we will need to sacrifice much less of the "raw" key in order to perform error correction and privacy amplification. Further quantitative data related to this is available upon request.

The QuNET system handles the distribution of the entangled states between Alice and Bob in the same way as it handles distribution for authentication protocols – requiring only a higher volume of QuSTICKS. As with the authentication protocols, the backbone QuNET system can be used to distribute and SWAP entanglement to Alice and Bob through a more complex networking structure if necessary.

The key distribution system using Turing's QuNET has extraordinary flexibility and significant benefits over a more traditional quantum key distribution network. As noted previously, the portability of the QuSTICKs allow QKD nodes to be placed essentially anywhere that a QuSTICK can be physically transported. Optic fibre, satellites and dedicated receiving stations, and other infrastructure-intensive hardware is not required. This allows for the distribution of QKD nodes to mobile platforms, field units, remote outposts, or forward operating bases. Additionally the entanglement network can be reconfigured as required. If QKD nodes need to be relocated from a mobile outpost in central Africa to the Middle east, we simply move the QuSTICKs using trucks or planes to redistribute the quantum hardware. We do not need to rebuild optic fibre links, re-position or re-launch costly satellite systems or rebuild complex optical receiving stations.

Key exchange in the QuNET environment essentially mimics the trusted courier model in which hard drives full of sensitive key material are physically transported around the world by trusted couriers. The security of classical keys are completely dependent on the reliability of these couriers. While vetting protocols used by the most skilled armed forces and intelligence agencies worldwide are generally excellent, the possibility of key material being lost, stolen, sold, or surreptitiously copied is the most significant failure mode of these networks.

In the QuNET system, *entanglement is the only thing being transported.* The keys themselves are not generated until just before being used. The QuSTICKs are programmed to generate the keys, use those keys for encryption, and then destroy the keys internally within the QuSTICK units. This allows us to generate and purify keys milliseconds before they are used to encrypt a classical message, then destroy them immediately thereafter. This does not completely close the window in which a key can be intercepted or copied, but it reduces it from hours or days (the time needed for a trusted courier to transport a hard drive from home base to a field unit) to mere milliseconds. Additionally, the only way in which keys could be copied or otherwise intercepted is if an individual physically steals or otherwise takes control of the QuSTICK itself. If an adversary does compromise a QuSTICK unit, they would also need to compromise the classical authentication channel that is used for key reconciliation and the transmission of the encrypted message. At the same time, this adversary would have to ensure that whatever physical act was performed in order to gain physical control over the QuSTICK had not been detected (refer back to the note above about hitting the intelligence agent over the head and stealing his or her briefcase).

6.3 One-time-pads

With the QuNET system, generation and usage of one-time-pad material occurs in exactly the same way as for key exchange. The only difference is the physical volume of QuSTICKs necessary. For key exchange, we generate a small, fixed amount of key material that is then combined with

strong, classical encryption techniques to generate a secure message. This opens up a possible security flaw because we have to trust the security of this classical encryption protocol (e.g. AES) at a time when quantum computers are on the verge of becoming practical. For highly sensitive information that needs to be secured for a long period of time, hoping that classical encryption techniques will remain secure for decades to come is an assumption that many governments and other entities may not wish to make.

One-time-pads are the only encryption technique discovered that is completely secure when implemented correctly. However, as discussed above, one-time-pad encryption requires a key that is exactly the same length as the message to be transmitted. The key material is hashed with the message and, provided that the key is only known by the sender and receiver, it is provably impossible for any eavesdropper to decrypt the message.

At the extreme end of the scale, a high-definition video stream (4K at 24 frames per second) requires transmitting approximately 35 MBps of classical data. Hence, securing this kind of transmission using a one-time-pad would require at least 70 million QMU's for each second of video. This is clearly an extraordinarily large number of Turing QuSTICKS. However, as we have noted, the QuNET system continues to grow as more physical QuSTICKS are produced.

In Fig. 12 we illustrate this assuming Moore's law scaling in the total number of QMU's fabricated (note that Moore's law actually quantifies the number of transistors that can be placed on a single chip; the numbers here are much, much larger). Fig. 12 shows the total number of transistors manufactured per year, worldwide, starting with the first demonstration of the transistor in 1947, until now (source: VSLI research, http://www.vsliresearch.com).

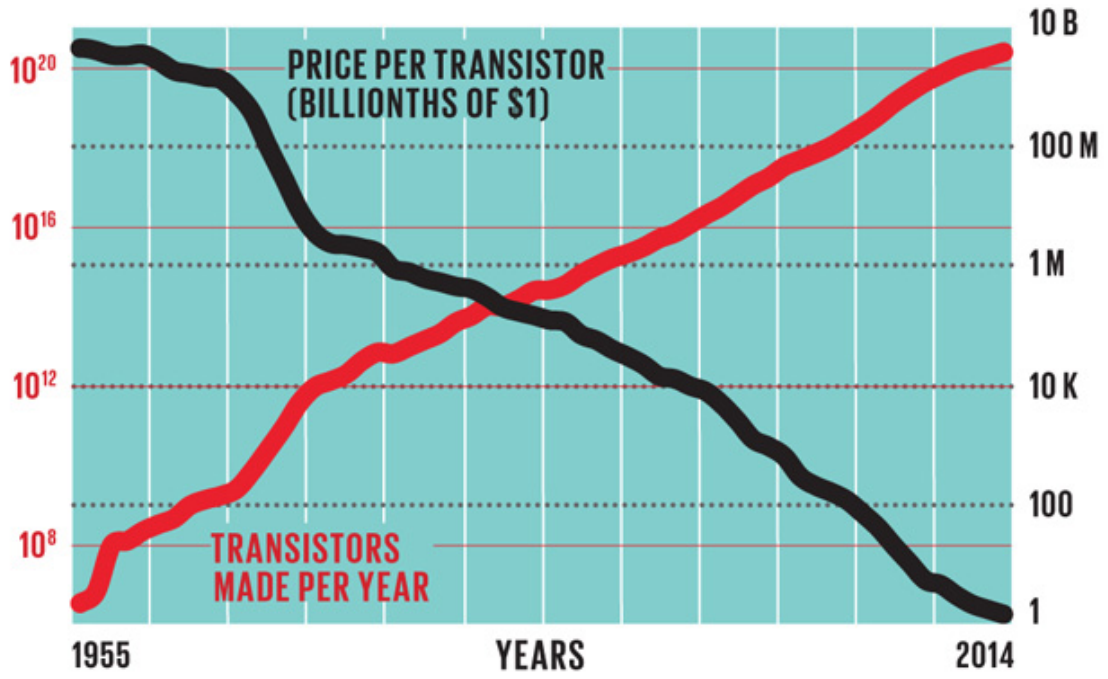

Figure 12: Historical number of transistors manufactured per year since their invention (red curve). (source: VSLI research, http://www.vsliresearch.com).

Assuming an initial R&D time frame of approximately 5 years for the first demonstration of a QMU, we further assume the same scaling in terms of the number of QMU's manufactured and illustrate the growth of the QuNET though each state: 1) a large-scale, global authentication network, 2) a global key exchange network, and 3) a global one-time-pad encryption network. We also show the forth and fifth stages of the QuNET: A classical information exchange through quantum entanglement (described in the next section), and, finally, the ability to build a true Quantum Internet that connects large-scale, error-corrected quantum computing systems. Fig. 13 illustrates

network volume extrapolated using the same scaling as Fig. 12 (in terms of the total number of QMU's manufactured after 2023) for the next 50 years.

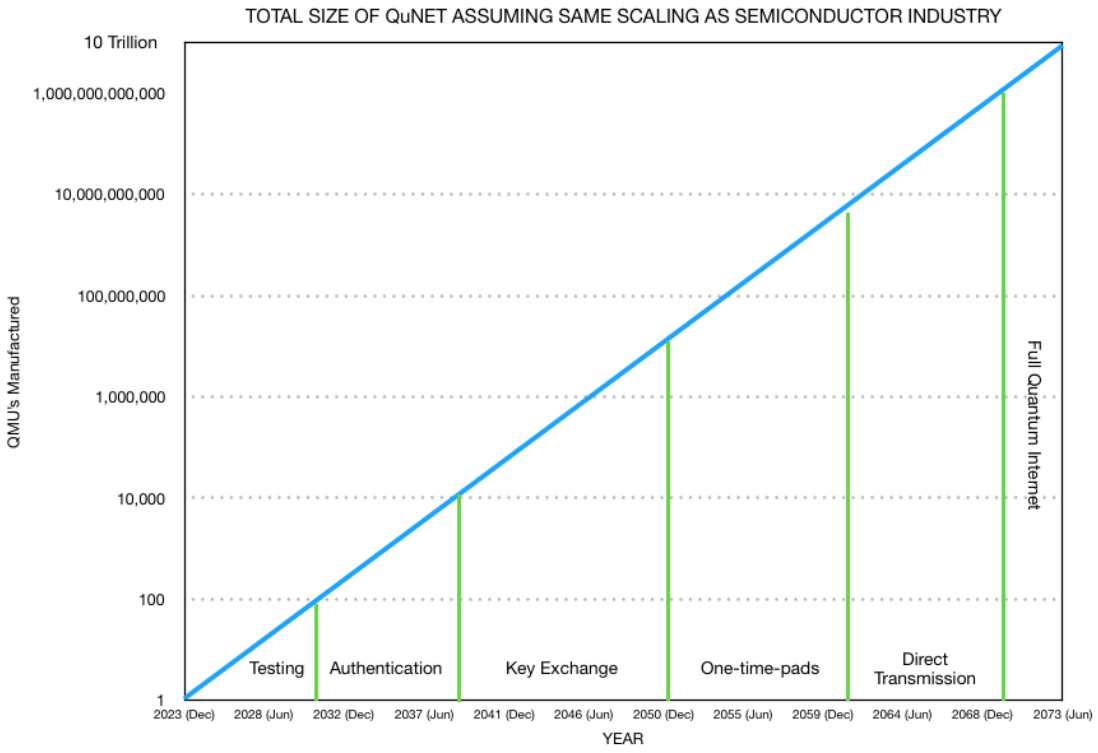

Figure 13: Assuming the same scaling as Figure 12 for the quantum world, the cumulative number of QMU's in existence for the next 50 years. This plot assumes that the first QMU is demonstrated in 2023 and a Moore's law scaling kicks in at that point. Illustrated are the boundaries when certain quantum protocols "come online", from small scale testing all the way to a functional, global quantum internet.

If a similar trend line is followed, approximately ten trillion QMU's can be manufactured by 2073. This is an astronomical figure for active quantum devices; and it would also be more than sufficient to construct multiple, large-scale quantum computing systems capable of executing any large-scale algorithm that has ever been developed. But keeping our discussion limited to the communications space, the QuNET system is, as we have said, designed to incorporate new QMU's as they become available, rather than replacing them. So the total size of the QuNET system is cumulative.

How such a network will be used is less clear. As noted above, we can make the following arguments as to when a particular application "comes online"

- Authentication tokens can come online once the number of QMU's > 100.
- Key exchange can come online once the number of QMU's $> 10,000$
- One-time-pads can come online once the number of QMU's $>$ 50Million. This assumes voice communications using the 600 bps NATO standard STANAG-4591 vocoding technique with a cumulative total of 10 hours of communication that needs to be encoded using QuSTICKS https://www.public.navy.mil/jtnc/APIs/API_1.4_20150226_VocoderService.pdf.
- Secure direct communications of classical information using quantum channels once the number of QMU's $>$ 1 Billion.
- Direct coupling of fully error-corrected quantum computers at a MHz logic-gate rate once the number of QMU's $>$ 10 Billion.
- Large-scale quantum internet once the number of QMU's $>$ 1 trillion.

Between 100 and 10,000 QMU's, the QuNET can expand into a network consisting of 100 separate authentication links, with each link implemented at arbitrary distance scales. These separate links could be deployed between the same two parties (increasing the speed of a single link) or it could be spread out over multiple locations. How the network is configured and re-configured is completely at the discretion of whoever owns or controls the physical QuSTICKS.

Once a total of 10,000 QMU's have been manufactured, the operator of the QuNET may choose to re-task *all* the QMU's then deployed to form a single link dedicated to key exchange for symmetric encryption protocols such as AES. Once this is achieved, the process starts again. At this point the network can become a hybrid of both key exchange and authentication, with resources deployed depending on need. The expansion process continues using these two protocols until one-time-pad quantum encryption becomes feasible for data transmission of significant volume (at approximately 50 Million QMU's). At each stage of the QuNET expansion, resources can be dynamically redesigned to support whatever protocols are both possible (given the total number of units in the field) and what is desired by whomever controls the physical QuSTICK units.

Once the number of QMU's exceeds the number needed to connect together fully error-corrected quantum computing systems, the network will continue to expand to encapsulate all the subprotocols, completing the transition to a multi-purpose Quantum Internet.

Our extrapolations based upon the historical evolution of the classical semiconductor industry may be optimistic; but given the design of Turing's technology, once a single QMU chip-set can be fabricated at low enough cost (we initially are targeting $1 per physical qubit, approximately $1000 per QMU, as mentioned earlier), expansion of the network becomes purely a function of how quickly high volume manufacturing can be developed and how much per-qubit costs can be reduced as that manufacturing infrastructure becomes more advanced.

6.4 Direct Information transfer

The last stage of the communications stack that can be achieved with the Turing QuNET before classifying the system as a general purpose Quantum Internet is the direct transmission of classical information using quantum channels. This may or may not make sense to do in practice – many would argue that access to a network capable of generating quantum one-time-pads is sufficient to perform any task in the communications space other than connecting together actual quantum computers.

New protocols become viable once high-bandwidth and high-fidelity QuNET-based communications channels exist. In Appendix B we consider the classical transmission of data via a quantum channel which can either be encrypted or not using the intrinsic properties of quantum mechanics. Unlike other protocols, we use the quantum channel directly to transmit a classical message rather than using the channel to provide a shared key to encrypt an otherwise standard classical data stream. The specific mathematical details are detailed in Appendix B, where we demonstrate that, to transmit an 8-bit classical message (with or without encryption added) using a rudimentary continuous variable encoding scheme, approximately 26,000 QMU's are required.

7 Responding to GCHQ and NSA criticisms of QKD

Now that we have described the structure and operation of the Turing QuNET system, let us revisit some of the criticisms of the GCHQ (UK's Government Communications Headquarters office) and NSA (USA's National Security Agency) mentioned earlier in using Quantum Key Distribution (QKD) to secure classical communications channels. GCHQ's public statement can be found here on NCSC.gov.uk. NSA's public statement can be found here on NSA.gov.

QKD protocols address only the problem of agreeing on keys for encrypting data. Ubiquitous on-demand modern services, such as verifying identities and data integrity, establishing network sessions, providing access control, and automatic software updates, rely on authentication and integrity mechanisms (e.g. digital signatures) as well as encryption.

As we have discussed, key distribution is only one protocol in a larger stack of applications of the QuNET system. The QuNET can be used for authentication and session integrity by using shared entanglement between two parties to violate Bell inequalities; and the portability of the QuNET system allows for these authentication channels to be set up wherever physical access to QuSTICK transport is available.

7.1 Criticisms from both GHCQ and NSA

1) The two major functional limitations of commercial QKD systems are the relatively short effective range of transmission, and the fact that BB84 and similar proposals are fundamentally point-to-point protocols. This means that QKD does not integrate easily with the Internet or with the mobile technologies, apps and services that dominate public and business life today.

Turing's QuNET system is designed from the ground up to be functional over global distances. It doesn't need extensive infrastructure along the communications channel and can leverage classical cargo transport mechanisms that are already available.

While we have illustrated much of the function of the Turing QuNET system through point-to-point protocols between two parties, the QuNET itself is a fully-formed quantum network. Entanglement can be distributed and shared between an arbitrary number of parties and more complex protocols, such as secret sharing or distributed communication/computation, can be performed. In the context of QKD, we do not utilize the point-to-point protocol of BB84. Instead we base QKD protocols on the more advanced Ekert-91 (which utilizes two properties of entanglement, the perfect correlation of measurements made by Alice and Bob as well as the ability to detect eavesdropping by noting disturbances in the quality of that correlation). QuNET's ability to distribute, share and maintain entanglement on global distances makes Ekert-91 practical for real-world cryptography for the first time.

2) Hardware is expensive to obtain and maintain. Unlike software, hardware cannot be patched remotely or cheaply when it degrades or when vulnerabilities are discovered.

As discussed above, the Turing QuNET is built from the QMU chip sets and the QuSTICK devices. These devices are designed to be cheap and to be mass manufactured, with network capabilities increasing as more and more units are produced. One of the major benefits to the QuNET system is that we can easily replace/repair or augment QuSTICKS within the network to fix potential vulnerabilities in the future or to simply fix faulty units. Unlike infrastructure-intensive quantum communication systems (such as quantum repeater networks or satellites), we have trivial access to each physical device within the network, and they can be repaired or replaced as simply as a non-functional hard-drive in a classical sneakernet communications channel. Patching hardware remotely is not required, as the hardware itself is easily portable. Hence, if repairs or patches are needed, QuSTICK units can be rotated in and out of the larger network and then immediately redeployed in the field without significantly affecting the performance of the network. Once large volumes of QuSTICKs are deployed in the network, regular servicing and repairs of individual units will be background noise to overall network performance.

3) Any real-world QKD system will be built from classical components, such as sources, detectors and fibers, and potentially ancillary classical network devices, any one of which may prove to be a weak link. A number of attacks have been proposed and demonstrated on deployed QKD systems that subvert one of more of these hardware components, enabling the secret shared key to be recovered without triggering an alarm.

Sources, detectors and fibres are not components of the Turing QuNET system. The integrity of the network rests with the functional integrity of individual QuSTICKs. As these units will be continuously moved, entangled locally, and then moved again, QuSTICKs can be tested and verified when the entanglement is initially prepared. Various unit testing can be performed be-

tween two QuSTICKs at home base to ensure quantum integrity of the system before they are ever deployed in the field. Compromised units can be removed from the network or returned for a complete rebuild and redeployment. Any units that become compromised in the field will be detected through the entanglement links that were initially prepared when QuSTICKs were present at home base. As no further entanglement operations are performed between QuSTICK units *after* they have been locally entangled and verified, points of failure (where security could be affected) are drastically reduced.

4) Denial of service (DoS) attacks that interfere with the paths carrying the QKD transmissions also seem potentially easier with QKD than with contemporary Internet or mobile network technologies. Since QKD devices typically abort a key establishment session when they detect tampering, this makes it difficult to recommend QKD for contexts where DoS attacks are likely to be attempted.

As we have mentioned, denial of service attacks require an actual physical QuSTICK to be compromised. For an effective denial of service attack to be launched against the entire network, an adversary would have to steal or physically compromise *every* QuSTICK unit one of the parties possesses – a much more difficult thing to do than simply cutting an optic fibre link or jamming the transmission of photons. If only a subset of QuSTICKS are stolen or otherwise compromised, network performance will decrease, but provided there are at least two uncompromised units somewhere, a viable entanglement connection will exist.

7.2 Criticisms from NSA

1) QKD generates keying material for an encryption algorithm but is not useful without the assurance that the original QKD transmission comes from the desired entity. QKD does not provide a means to authenticate the QKD transmission source.

Quantum memory unit based communication relies on the physical transport of one of two quantum hard drive systems from sender to receiver. This transport alone provides a certain level of verified security as, initially, quantum hard drives will be somewhat cumbersome hardware that will not go unnoticed if it is physically stolen or substituted with another unit. However, there are entanglement based solutions to providing secure authentication between sender and receiver.

In summary, the physical building blocks of the Turing QuNET effectively negate all the major concerns arising from GCHQ's and NSA's assessments of quantum technology for encryption and security. We argue that the flexibility of this system makes it the only viable method currently on the table for quantum communications protocols, including QKD, network authentication, as well as secure classical communications.

8 A Vision for the Future: Creating a QuNET Quantum Internet

Utilizing the QuNET to directly transmit (either encrypted or unencrypted) classical data is more resource intensive than authentication protocols, key exchange, or, arguably, one-time-pad encryption data. However, there may be cases where having this flexibility in the QuNET would be beneficial. As shown in Fig.13, approximately 30,000 QMU's are required to transmit a single 8-bit string of classical information with high accuracy, though this can be reduced by using more advanced encoding techniques for the classical message. 30,000 QMU's are sufficient to perform 300 authentication transactions or generate fifteen to twenty 256-bit encryption keys for symmetric AES protocols.

Such high data-rate classical communications utilizing the QuNET's quantum channels are the last stage before a full quantum Internet is possible. The physical QMU resources needed to connect together fully error-corrected quantum computing systems would only be about an order of magnitude higher than using quantum methods to directly transmit encrypted or unencrypted classical data. A fully functional Quantum Internet (that is, a network with the ability to provide quantum

communication channels between large-scale, error corrected quantum computers [DMN11, Fit17]) requires the following:

- Transcontinental communication links spanning distances anywhere up to 10,000Km
- A high speed network spanning those distances for $>$ THz operational speeds.
- End-to-end error rates of 10^{-10} or lower.

Two ideas for constructing such worldwide quantum communication networks have received extensive theoretical examination and experimental demonstration: 1. Quantum repeater systems. 2. Quantum-based satellite communications. However, both approaches run into significant practical limitations.

High-speed quantum repeater networks only exist for *theoretical* transmission rates of about 1-10 MHz [FWH+10] and require repeater stations every 20 to 50 km [MSD+12]. This upper limit to repeater station separation is necessitated by the loss rates of current optic fibre technology. It is unclear if that range can ever be extended. An associated problem with the small separation distances of quantum repeater stations is the difficulty of deploying networks across oceans or otherwise inhospitable environments. Quantum repeaters are small-scale quantum computers, consisting of a few thousand qubits and associated control infrastructure. This requirement currently precludes their deployment at high densities across the planet.

Regarding satellite technology, there has been significant experimental progress, with entanglement-based satellite platforms deployed by the Chinese, as well as proof-of-principle payloads deployed by the Singaporeans, Japanese and Austrians. These platforms are not designed for general purpose quantum communications. They are built for QKD applications, which do not have the stringent constraints listed above.

While fully error-corrected, high bandwidth, and low error-rate satellite systems have not received significant theoretical attention as yet, we can safely assume that such technology *could* be built and deployed. However, deployment and maintenance costs, bandwidth sufficient for fully error-corrected communications channels, and the infrastructure associated with receiving stations represent very big hurdles to overcome when utilizing space based platforms as backbones to a Quantum Internet.

QuSTICK-based quantum networks can satisfy the constraints noted above as well as address the issues associated with quantum repeaters and satellite communication systems. Initial analysis shows that $>$ THz transpacific networks can be realised in a system such as the Turing QuSTICK, provided that the cost per physical qubit is low enough. Beyond the challenge of simply building a sufficient number of QuSTICKs (a challenge that is common for any hardware company targeting large-scale quantum computing platforms), there is no additional hardware development work needed to realise a global network. The only additional infrastructure needed by a global QuSTICK network is traditional global shipping channels that already exist.

Prototype quantum communication networks using sneakernet principles and QuSTICKs will first be demonstrated at much shorter ranges and at slower communication rates. Scaling up to higher fidelity, higher speed, and longer ranges is conceptually straightforward. Additionally, communications channels do not influence infrastructure development. Building a quantum communications system between Tokyo and Osaka is no different to building a communications system from Japan to Australia.

8.1 Flexible modes of operations

One of the most exciting aspects of the Turing QuSTICK hardware platform is the ability to program in multiple modes of operation without having to redesign or rebuild the hardware itself. The optically mediated coupling between NV-defect qubits allows us to go far beyond a standard 2D nearest neighbor geometric constraints that are associated with other quantum hardware architectures.

While error correction techniques such as surface code are now preferred models for systems such as ion traps and superconductors, it is not necessarily the best choice for every potential application of a mid to large-scale qubit array. In the Turing QuSTICK, we can reconfigure, on the fly, the optical pathways that connect our array of NV qubits. This permits us to exploit other error correction techniques that may require connection geometries that deviate from 2D nearest-neighbor. As these optical reconfigurations do not change how we operate the NV qubits themselves, this reconfiguration can be performed on the fly to create qubit connection geometries that correspond to the most appropriate error correction techniques for a specific application.

Let's look at examples of three possible modes of operation:

- In the first, no error correction is employed and the QuSTICK acts as a small to mid-scale quantum computer operating without error correction. Commercial applications in the regime of non-error corrected qubits do not yet exist; but this mode of operation could be used to demonstrate a quantum-supremacy protocol (as a way to test the integrity of a QuSTICK) or to provide a testing and/or training platform for quantum application developers.
- The second example is using the QuSTICK as an entanglement distribution platform. Utilizing each chip within the QuSTICK as a quantum memory may require a more efficient quantum error correction code to increase the density of *logical* qubits in the machine. Since entanglement distribution does not require an error-corrected, universal gate set, other coding choices with finite rate codes or better logical error rate performance may be desirable.
- The third example is fully fault-tolerant quantum computation, where error correction codes that are most efficient for universal quantum computation are desired.

In each case, reconfiguration would occur with pre-loaded Quantum BIOS profiles that are directly loaded into the Turing QuSTICK. The user would not be expected to choose (or even know about the details of) these configurations, but instead the Quantum BIOS would be responsible for analyzing the details of the desired application and then configuring the QuBE into the most appropriate operational mode.

By having the flexibility to run the QuSTiCK in multiple different modes, depending on the specifics of the problem, we can also take advantage of any further theoretical developments in error correction, even if they occur after the construction of the Turing QuSTiCK. Rather than redesigning and rebuilding the qubit array if more powerful error correction techniques are developed, we instead provide software updates to the Quantum BIOS that enables a new mode of operation based upon new techniques developed, potentially long after units are built, sold, and deployed.

References

[AGA+18] T. Astner, J. Gugler, A. Angerer, S. Wald, S. Putz, N. J. Mauser, M. Trupke, H. Sumiya, S. Onoda, J. Isoya, J. Schmiedmayer, P. Mohn, and J. Majer. Solid-state electron spin lifetime limited by phononic vacuum modes. *Nature Materials*, 17:313, 2018.

[AGP11] Igor Aharonovich, Andrew D. Greentree, and Steven Prawer. Diamond photonics. *Nature Photonics*, 5:397, 2011.

[ATL15] Koji Azuma, Kiyoshi Tamaki, and Hoi-Kwong Lo. All-photonic quantum repeaters. *Nature Communications*, 6:6787, 2015.

[BDR00] E.R. Berndt, E.R. Dulberger, and N.J. Rappaport. Price and quality of desktop and mobile personal computers: A quarter century of history. *Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association*, 91:268, 2000.

[Bom15] Héctor Bombín. Gauge color codes: optimal transversal gates and gauge fixing in topological stabilizer codes. *New Journal of Physics*, 17:083002, 2015.

[BVC+17] Nikolas P Breuckmann, Christophe Vuillot, Earl Campbell, Anirudh Krishna, and Barbara M Terhal. Hyperbolic and semi-hyperbolic surface codes for quantum storage. *Quantum Science and Technology*, 2:035007, 2017.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880, 1969.

[CS01] Marcos Curty and David J. Santos. Quantum authentication of classical messages. *Physical Review A*, 64:062309, 2001.

[DGSVM16] Simon J. Devitt, Andrew D. Greentree, Ashley M. Stephens, and Rodney Van Meter. High-speed quantum networking by ship. *Scientific Reports*, 6:36163, 2016.

[DMN11] S.J. Devitt, W.J. Munro, and K. Nemoto. High Performance Quantum Computing. *Progress in Informatics*, 8:49, 2011.

[DMN13] Simon J Devitt, William J Munro, and Kae Nemoto. Quantum error correction for beginners. *Reports on Progress in Physics*, 76:076001, 2013.

[FGL18] O. Fawzi, A. Grospellier, and A. Leverrier. Constant overhead quantum fault-tolerance with quantum expander codes. *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, page 743, 2018.

[Fit17] Joseph F. Fitzsimons. Private quantum computation: an introduction to blind quantum computing and related protocols. *NPJ Quantum Information*, 3:23, 2017.

[FMMC12] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86:032324, 2012.

[FWH+10] Austin G. Fowler, David S. Wang, Charles D. Hill, Thaddeus D. Ladd, Rodney Van Meter, and Lloyd C. L. Hollenberg. Surface code quantum communication. *Physical Review Letters*, 104:180503, 2010.

[GRW+07] R.D. Gehrz, T.L. Roellig, M.W. Werner, G.G. Fazio, J.R. Houck, F.J. Low, G.H. Rieke, B.T. Soifer, D.A. Levine, and E.A. Romana. The NASA Spitzer Space Telescope. *Review of Scientific Instruments*, 78:011302, 2007.

[HFDM12] Clare Horsman, Austin G Fowler, Simon Devitt, and Rodney Van Meter. Surface code quantum computing by lattice surgery. *New Journal of Physics*, 14:123011, 2012.

[Mer78] Ralph C. Merkle. Secure communications over insecure channels. *Communications ACM*, 21:294, 1978.

[MHS+12] Xiao-Song Ma, Thomas Herbst, Thomas Scheidl, Daqing Wang, Sebastian Kropatschek, William Naylor, Bernhard Wittmann, Alexandra Mech, Johannes Kofler, Elena Anisimova, Vadim Makarov, Thomas Jennewein, Rupert Ursin, and Anton Zeilinger. Quantum teleportation over 143 kilometres using active feed-forward. *Nature*, 489:269, 2012.

[Mil67] S. Milgram. The small world problem. *Psychology Today*, 2:60, 1967.

[MSD+12] W. J. Munro, A. M. Stephens, S. J. Devitt, K. A. Harrison, and Kae Nemoto. Quantum communication without the necessity of quantum memories. *Nature Photonics*, 6:777, 2012.

[Pre18] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.

[RHAS17] Miloš Rančić, Morgan P. Hedges, Rose L. Ahlefeldt, and Matthew J. Sellars. Coherence time of over a second in a telecom-compatible quantum memory storage material. *Nature Physics*, 14:50, 2017.

[Sha49] C. E. Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28:656, 1949.

[SSdRG11] Nicolas Sangouard, Christoph Simon, Hugues de Riedmatten, and Nicolas Gisin. Quantum repeaters based on atomic ensembles and linear optics. *Reviews of Modern Physics*, 83:33, 2011.

[TBF18] David K. Tuckett, Stephen D. Bartlett, and Steven T. Flammia. Ultrahigh error threshold for surface codes with biased noise. *Physical Review Letters*, 120:050505, 2018.

[TCCF+17] Hideki Takenaka, Alberto Carrasco-Casado, Mikio Fujiwara, Mitsuo Kitamura, Masahide Sasaki, and Morio Toyoshima. Satellite-to-ground quantum-limited communication using a 50-kg-class microsatellite. *Nature Photonics*, 11:502, 2017.

[TCT+16] Zhongkan Tang, Rakhitha Chandrasekara, Yue Chuan Tan, Cliff Cheng, Luo Sha, Goh Cher Hiang, Daniel K. L. Oi, and Alexander Ling. Generation and analysis of correlated pairs of photons aboard a nanosatellite. *Physical Review Applied*, 5:054022, 2016.

[Ter15] Barbara M. Terhal. Quantum error correction for quantum memories. *Reviews of Modern Physics*, 87:307, 2015.

[UTSM+07] R. Ursin, F. Tiefenbacher, T. Schmitt-Manderbach, H. Weier, T. Scheidl, M. Lindenthal, B. Blauensteiner, T. Jennewein, J. Perdigues, P. Trojek, B. Ömer, M. Fürst, M. Meyenburg, J. Rarity, Z. Sodnik, C. Barbieri, H. Weinfurter, and A. Zeilinger. Entanglement-based quantum communication over 144 km. *Nature Physics*, 3:481, 2007.

[WEH18] Stephanie Wehner, David Elkouss, and Ronald Hanson. Quantum internet: A vision for the road ahead. *Science*, 362:eaam9288, 2018.

[YCL+17] Juan Yin, Yuan Cao, Yu-Huai Li, Sheng-Kai Liao, Liang Zhang, Ji-Gang Ren, Wen-Qi Cai, Wei-Yue Liu, Bo Li, Hui Dai, Guang-Bing Li, Qi-Ming Lu, Yun-Hong Gong, Yu Xu, Shuang-Lin Li, Feng-Zhi Li, Ya-Yun Yin, Zi-Qing Jiang, Ming Li, Jian-Jun Jia, Ge Ren, Dong He, Yi-Lin Zhou, Xiao-Xiang Zhang, Na Wang, Xiang Chang, Zhen-Cai Zhu, Nai-Le Liu, Yu-Ao Chen, Chao-Yang Lu, Rong Shu, Cheng-Zhi Peng, Jian-Yu Wang, and Jian-Wei Pan. Satellite-based entanglement distribution over 1200 kilometers. *Science*, 356:1140, 2017.

[ZHA+15] Manjin Zhong, Morgan P. Hedges, Rose L. Ahlefeldt, John G. Bartholomew, Sarah E. Beavan, Sven M. Wittig, Jevon J. Longdell, and Matthew J. Sellars. Optically addressable nuclear spins in a solid with a six-hour coherence time. *Nature*, 517:177 EP, 2015.

[ŻZHE93] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert. "event-ready-detectors" bell experiment via entanglement swapping. *Physical Review Letters*, 71:4287, 1993.

A Appendix - Authentication Protocols

Alice and Bob are sharing a *logical* Bell state of the form

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right) \tag{1}$$

The subscripts are the *logical qubits* associated with (A)lice and (B)ob respectively and each logical qubit is stored within a single QMU within each of their respective QuSTICKs.

Alice and Bob then decide to authenticate using a Bell violation. This requires multiple copies of the above state (each contained within a separate QMU of the QuSTICK), entangled between Alice and Bob. To perform a Bell violation, Alice and Bob randomly choose a set of "settings" to measure each logical qubit in their QuSTICKS. These measurement "settings" are performed by Alice randomly choosing the following sets of single qubit rotations just prior to measuring her half of the logical Bell state in the $|0\rangle, |1\rangle$ basis:

$$\begin{aligned} U_A^1 &= I \quad \text{with probability } p = 0.5 \\ U_A^2 &= e^{-i\frac{\pi}{4}\sigma_x} \quad \text{with } p = 0.5 \end{aligned} \tag{2}$$

and Bob performing the following rotations just prior to his measurements in the $\{|0\rangle, |1\rangle\}$ basis:

$$\begin{aligned} U_B^1 &= e^{-i\frac{\pi}{8}\sigma_x} \quad \text{with } p = 0.5 \\ U_B^2 &= e^{-i\frac{3\pi}{8}\sigma_x} \quad \text{with } p = 0.5 \end{aligned} \tag{3}$$

Each of these rotations for both Alice and Bob are chosen independently and randomly (by, for example, Alice and Bob tossing a coin to choose what rotations to apply just prior to measurement).

For each measurement of the state, Alice and Bob will each obtain a classical bit value $\{s_{i,j}^A, s_{i,j}^B\} \in (0, 1)$. If we define the following, $E_{ij}(U_i^A, U_j^B) = \langle U_i^A U_j^B |\psi\rangle_{AB}\rangle$ as the expectation value of these classical bit values, i.e.

$$E_{i,j}(U_i^A, U_j^B) = \left(\frac{1}{N}\sum_{k=1}^{N} s_{i,j}^A(k)\right)\left(\frac{1}{N}\sum_{k=1}^{N} s^B(k)_{i,j}\right) \tag{4}$$

where we take the product of the average of Alice and Bob's classical measurement results over N copies of the Bell state, then quantum mechanics dictates that,

$$|E_{2,2} + E_{2,1} + E_{1,2} - E_{1,1}| = 2\sqrt{2} = S \tag{5}$$

if and only if they actually share an set of entangled states. If Alice and Bob are not sharing entanglement then $S \leq 2$. This is known as the Bell violation. If S is measured to be > 2, this confirms that Alice and Bob are actually sharing entanglement.

In order to calculate Eq. 5 Alice and Bob must have randomly applied the correct set of unitary gates prior to measurement. For example, to calculate $E_{1,1}$ Alice must have randomly applied the gate U_1^A on her half of the same Bell pair as Bob applied the gate U_1^B. As Alice and Bob choose their gates randomly, the probability of this occurring is $p = (0.5)^2 = 0.25$. This is true for all the quantities, $E_{i,j}$. The actual protocol consists of the following steps:

- Alice and Bob begin with M *logically encoded* copies of the shared Bell state $|\psi\rangle_{AB}$, with each respective half of each state stored within a separate QMU of Alice and Bob's respective QuSTICKS.
- Alice measures each of her QMU logical qubits, applying the rotations U_1^A and U_2^A randomly before measurement with a probability of 0.5
- Bob does the same, independently for each of his QMU's
- Alice and Bob now possess a set of M *classical bits*, with each of the M classical bits corresponding to the measurement of a specific, indexed QMU.

- Alice and Bob simultaneously share with each other both the measurement setting they randomly chose (which gate they respectively applied, U_1 or U_2) and the classical bit value they measured. It is important that this "reveal" step happens simultaneously between Alice and Bob so that neither of them have the ability to cheat and change their data. We discuss this more below.

- For each pairwise state $|\psi\rangle_{AB}$ shared between Alice and Bob and the respective QMU indicies for Alice and Bob's QuSTICKs, the classical values are sorted into four "buckets", depending on the random rotations U_i^A and U_j^B chosen by Alice and Bob. For the initial M shared logical states, each "bucket" will contain, on average, $M/4$ classical bit values.

- Alice and Bob then compute the averages over the $M/4$ values in each bucket and then calculate the respective $E_{i,j}$ value.

- The four $E_{i,j}$ values are then used to calculate S.

For $M/4$ classical bits that Alice and Bob have in each "bucket" we can define the error associated with each $E_{i,j}$ and hence S as

$$\left(\frac{\Delta E_{i,j}}{E_{i,j}}\right)^2 = \left(\frac{\Delta\langle s_{i,j}^A\rangle}{\langle s_{i,j}^A\rangle}\right)^2 + \left(\frac{\Delta\langle s_{i,j}^B\rangle}{\langle s_{i,j}^B\rangle}\right)^2 \tag{6}$$

under the assumption of independence for Alice and Bob and the fact that fair sampling of $M/4$ variables gives $\Delta s_{i,j}^{A,B} = 2/\sqrt{M}$. This is the standard error term, and so we can bound $\Delta E_{i,j} \approx 1/\sqrt{M}$ and hence $\Delta S \approx 4/\sqrt{M}$.

Bit string	Corresponding floating point number
00000000	0
00000001	$\pi/(2^8)$
00000010	$2\pi/(2^8)$
00000011	$3\pi/(2^8)$
00000100	$4\pi/(2^8)$
...	...
11111111	$255\pi/(2^8)$

Table 1:

B Appendix - Direct Information Transfer

We first consider a bi-bipartite quantum state shared between two parties, Alice and Bob. The state is a maximally entangled Bell state of the form

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\right) \tag{7}$$

where the subscripts A and B denote the logical qubit held by Alice and Bob respectively. The distribution of this high-fidelity entangled state is discussed in the implementation section.

We assume for the moment that the shared state between Alice and Bob is perfect (i.e. has Fidelity = 1 and has no errors associated with it). The classical information that will be transmitted using this entanglement resource will be encoded using the relative phase of the entangled state above. Specifically, if Alice wishes to send some information to Bob, she will perform a Z-axis rotation on her qubit by some angle θ, i.e.

$$R_z(\theta)_A|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}\left(|0\rangle_A|0\rangle_B + e^{i\theta}|1\rangle_A|1\rangle_B\right) \tag{8}$$

The manner in which the classical information string is encoded into the variable θ can take many forms. As an example, we consider the case where the binary representation of the classical string is encoded with a sufficiently high-precision floating point number between 0 and π.

For a single byte of classical data (an 8-bit string) we can use the mapping shown in the table above. In principle, each of these conversions from binary to floating point will have an associated precision to discriminate actual values. We will return to this issue later.

Once the state has been encoded by Alice, she is free to measure her half of the encoded Bell state. She measures in the X-basis, which can result in two possible outcomes with a 50:50 probability for each.

$$\begin{aligned}
&\text{Alice measures: } \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)_A \equiv |+\rangle_A \\
&\text{Bob's resultant state: } \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\theta}|1\rangle\right)_B \\
&\text{Alice measures: } \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right)_A \equiv |-\rangle_A \\
&\text{Bob's resultant state: } \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i(\theta+\pi)}|1\rangle\right)_B
\end{aligned} \tag{9}$$

This measurement disentangles Alice and Bob's qubits, but the phase information (θ) that carries the information is now completely contained within Bob's qubit. The result of Alice's measurement results in a shift of the transmitted information from θ to $\theta + \pi$, which can be corrected for by Alice classically transmitting to Bob the measurement result she obtained (either the $|+\rangle$ or $|-\rangle$ result). If she measures a $|-\rangle$ state then Bob simply applies a Z-flip on his qubit, i.e.

$$Z\frac{1}{\sqrt{2}}\left(|0\rangle + e^{i(\theta+\pi)}|1\rangle\right)_B = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\theta}|1\rangle\right)_B \tag{10}$$

This ensures that Bob always will have the same state (with the encoded θ) regardless of Alice's measurement result.

How does Bob extract the information encoded in this state (i.e. measure θ)? If Bob simply performs a $\{|0\rangle,|1\rangle\}$ measurement (a Z-basis measurement) on his qubit, he will measure each result with a 50% probability and be unable to ascertain any information about θ. However, if he performs the same X-basis measurement as Alice, his probability of measuring the $|+\rangle$ or $|-\rangle$ state is given by:

$$\begin{aligned} Pr(|+\rangle) &= \frac{|1+e^{i\theta}|^2}{4} = \frac{1}{2} + \frac{\cos(\theta)}{2} \\ Pr(|-\rangle) &= \frac{|1-e^{i\theta}|^2}{4} = \frac{1}{2} - \frac{\cos(\theta)}{2} \end{aligned} \tag{11}$$

Hence the probability of measuring certain outputs for Bob contains the information encoded by Alice.

Unfortunately, since Bob's measurement is a binary result (either $|+\rangle$ or $|-\rangle$), a single copy of the state is insufficient to calculate θ. Instead, Alice and Bob must share multiple copies of the state in order to accurately calculate $Pr(|+\rangle)$ and $Pr(|-\rangle)$ and hence determine θ. How many copies are required?

The expectation value (or average value measured by Bob) over many copies of the state is given by:

$$Av = Pr(|+\rangle) - Pr(|-\rangle) = \cos(\theta) \tag{12}$$

Our ability to accurately reconstruct this number is dependent on the number of copies of the state we have. As each copy is independent (and assumed to be encoded with the same variable, θ), the error associated with our calculation of $\cos(\theta)$ scales as $1/\sqrt{N}$, where N is the number of copies of the shared encoded state we begin with. Since the value of the transmitted classical variable, θ is given by $\theta = \cos^{-1}(Av)$, the error associated with determining this value is given by:

$$\Delta\theta \approx \frac{1}{\sqrt{N}\sin(\theta)} \tag{13}$$

i.e it also decreases as $1/\sqrt{N}$.

In the above example, we considered a single byte of classical communication between Alice and Bob. In this case, we encode 256 possible binary values in $\theta \in [0, \pi)$. In order to faithfully distinguish between neighboring binary variables, we need the error associated with our measurement of θ to be $\Delta\theta < \pi/(2 * 256)$. This allows us to bound the number of copies needed between Alice and Bob as:

$$\frac{1}{\sqrt{N}} < \frac{\pi}{512}, \quad N > 26560. \tag{14}$$

This is a large number of copies to ensure that we can accurately resolve the binary value stored within θ. This number increases as we attempt to encode larger and larger numbers into the transmitted phase θ encoded using the shared state.

As noted at the beginning of this section, we chose a simple conversion of a binary variable to a floating point variable to transmit the message. This is not the most efficient encoding scheme that could be utilised. Compressing a larger digital string in a floating point number would allow for us to transmit classical data in a much more efficient manner, given the number of shared copies needed to faithfully reconstruct the transmitted variable, θ.

The protocol flow for sharing classical information via the quantum channel is illustrated in Figure 14.

B.1 adding encryption

Quantum communications channels have been of considerable interest due to the fact that, in principle, they can provide an informationally secure quantum communications channel for the

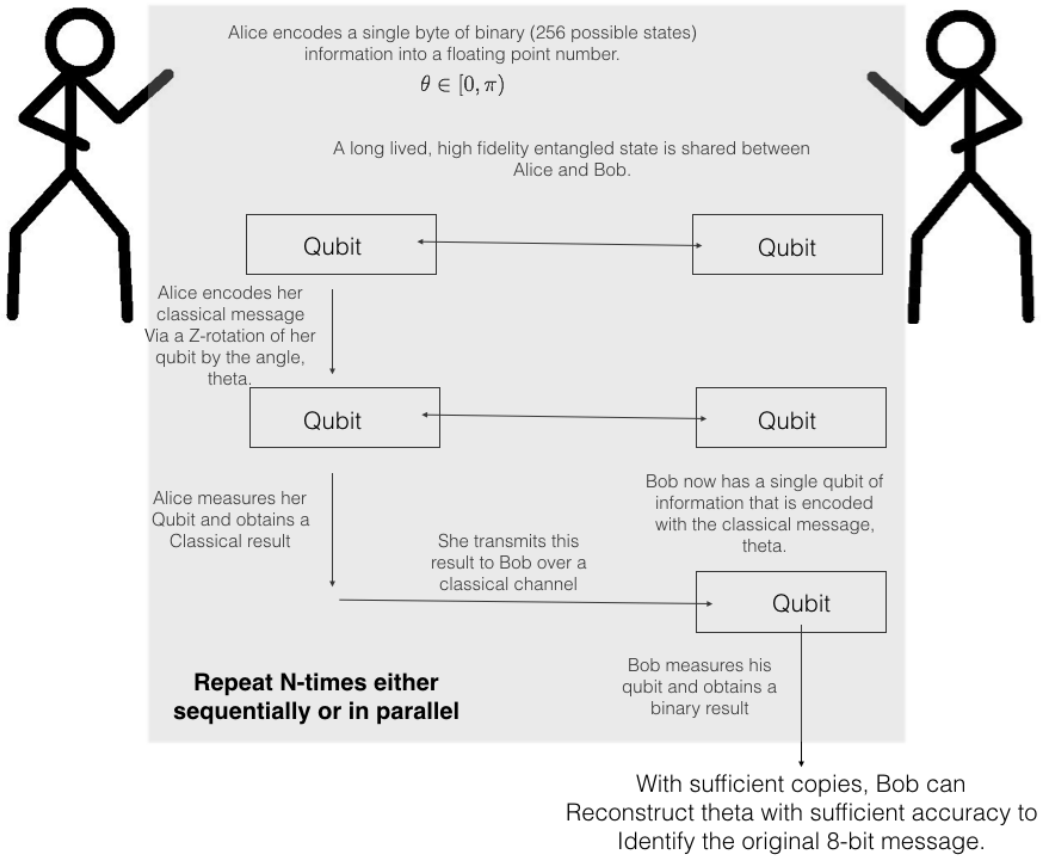

Figure 14: Protocol flow for sharing a classical message over a quantum communication channel.

production of a shared cryptographic key. By utilising quantum correlations contained within a shared entangled state, two parties (Alice and Bob) can establish a secure bit string that cannot be intercepted or contaminated by an eavesdropper without detection. In entanglement-based Quantum Key Distribution protocols (QKD), Alice and Bob share the entangled state:

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B\right) \tag{15}$$

Alice and Bob then independently choose to measure their respective qubits in either the Z-basis (i.e. do I have $|0\rangle$ or $|1\rangle$) or the X-basis (do I have $|+\rangle$ or $|-\rangle$). Provided that they both choose randomly to measure in the same basis (i.e. measure their respective qubits in the same way), they will share a bit stream that is always correlated. That is, if they both measure qubits in the same way, they will ALWAYS obtain the same results. This is the key insight that underlies the security of quantum information. In insures that this set of correlated results can be used to construct a secret key.

In order to construct the key from a random sequence of measurement choices for both Alice and Bob, the two parties publicly communicate which measurement choice they made. Given that they both independently can choose from one of two settings, 50% of the time, they will randomly make the same choice and 50% of the time they will make different choices. Any time they do not choose the same measurement setting they simply throw away the associated classical bit value they measure. The other 50% of the classical bits can then be used to form a secret key. Additional classical bits are utilised further from this shared key to perform functions such as error correction and eavesdropper detection, but these are subtleties to the protocol that are beyond the scope of this white paper.

In the protocol described in the previous section, we do not concern ourselves with the security of transmitting the information between Alice and Bob. We only want to use the quantum

channel to share a classical data string. However, we can also add an encryption layer to the protocol. After we share an entangled state between Alice and Bob, Alice encodes her information by performing a Z-rotation by an angle θ. However, she could choose to perform an X-rotation by the same angle. This transforms the shared state to:

$$R_x(\theta)|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}\left(|+\rangle_A|+\rangle_B + e^{i\theta}|-\rangle_A|-\rangle_B\right) \tag{16}$$

If Alice now measures her qubit in the Z-basis (and so does Bob), the same equations that we detailed in the previous section hold, and the variable θ can be extracted from the probability calculations obtained through multiple shared copies of the entangled state. As in the case shown in the previous section, if Bob does not measure in the Z-basis (and instead measures in the X-basis), he will obtain zero information about the state θ.

Consequently, in order to add an encryption layer to the communications channel, Alice takes roughly 50% of her shared copies and encodes the value θ via a $R_z(\theta)$ gate on her qubit and encodes the other 50% of her shared copies with value θ via an $R_x(\theta)$ gate on her qubit before performing measurements in the associated basis.

Her random encoding is then classically broadcast to Bob who has also randomly chosen to measure approximately 50% of his shared state in the Z-basis and the other 50% in the X-basis. For shared states where Alice and Bob's choices differ, the resultant classical information is simply thrown away. For the cases where Alice and Bob choose the same encoding/measurement basis, Bob will reconstruct the classical information θ in the same way as described in the previous section.

To maintain similar levels of security as standard QKD protocols, some of the shared states between Alice and Bob will be used for error correction and privacy amplification to ensure that interception of the encoded information is not possible. These protocols are standard for all QKD applications. The protocol flow for an encrypted sharing of classical information via the quantum channel is illustrated in Figure 15.

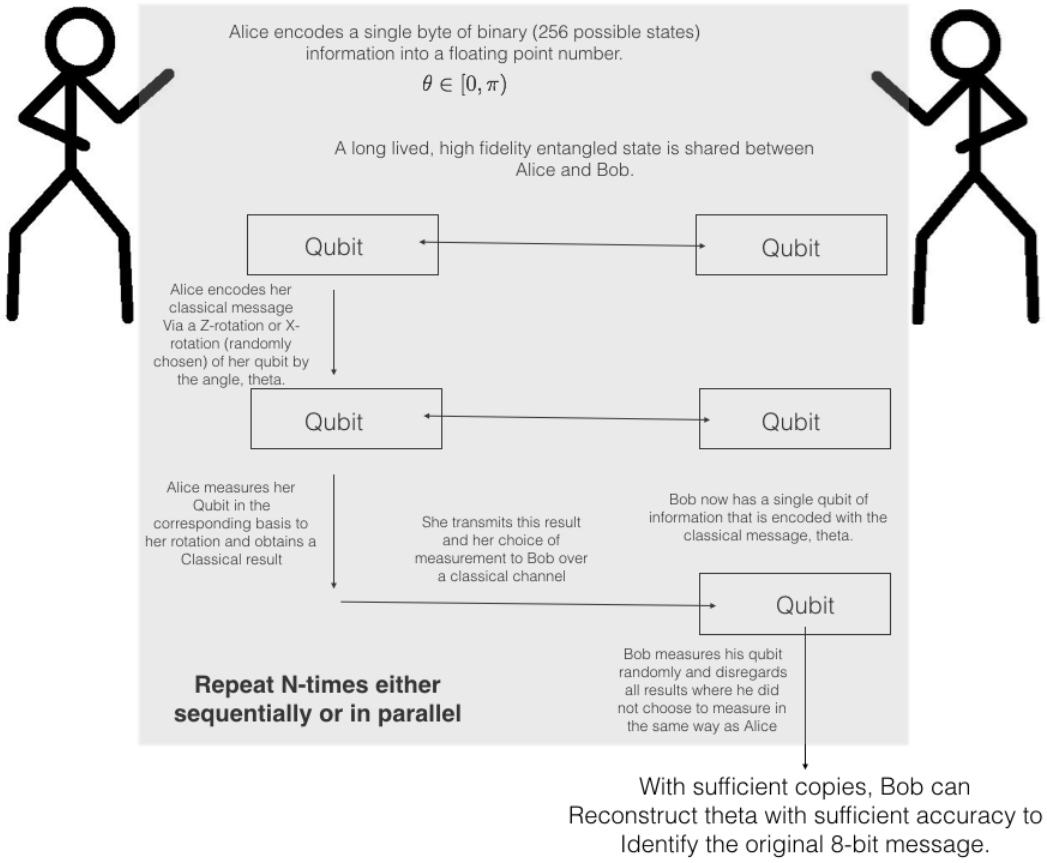

Figure 15: The protocol for embedding encryption into the communications protocol requires Alice to randomly choose 50% of her Bell states to encode using a $R_z(\theta)$ rotation or an $R_z(\theta)$ rotation and measure in the corresponding basis. Bob similarly measures his qubit in a random basis (either X or Z) and then disregards all classical results when his choice differs from Alice. He may then use some of the states for further cryptographic error correction or privacy amplification to protect against eavesdroppers and the remaining subset of qubit measurement results can then be used to reconstruct the classical message.