CLAIMS

WHAT IS CLAIMED IS:

1. A method for encoding data, the method comprising:

(a) splitting data into a plurality of frames, wherein each frame in the plurality of frames comprises a frame index;

(b) applying an outer codec to each frame in the plurality of frames, wherein the outer codec comprises an error correction scheme;

(c) dividing each frame into a plurality of lanes, wherein each lane in the plurality of lanes comprises a lane index;

(d) shuffling each lane based at least in part on the lane index; and

(e) applying an inner codec to encode each lane in a corresponding polynucleotide sequence, thereby generating a plurality of corresponding polynucleotide sequences.

2. The method of claim 1, wherein the data comprises binary data, including a byte stream or a byte array.

3. The method of claim 1, wherein the shuffling in (d) comprises a rotation scheme within each lane, comprises a pseudorandom process within each lane, and/or provides resistance against errors.

4. The method of claim 3, wherein the shuffling in (d) provides resistance against errors, and the errors are nucleotide synthesis errors or sequencing errors.

5. The method of claim 4, wherein the errors comprise a deletion, an insertion, or a substitution.

6. The method of claim 1, wherein the error correction scheme comprises a Reed-Solomon (RS) code, a low-density parity-check (LDPC) code, a Turbo-code, a polar code, or any combination thereof.

7. The method of claim 1, wherein:

the plurality of frames comprises about 100 to about 10,000 frames;

each frame comprises less than or equal to about 5000 lanes;

each lane comprises about 100 to about 300 bits;

each frame index comprises about 16 to about 20 bits;

the lane index comprises about 12 bits or about 16 bits; and/or

each corresponding polynucleotide sequence of the plurality of corresponding polynucleotide sequences has a length of about 100 to about 300 bases.

8. The method of claim 1, wherein the frame index and/or the lane index are prepended to each lane prior to (d).

9. The method of claim 1, wherein applying the inner codec comprises adding redundancy across the plurality of corresponding polynucleotide sequences, wherein the redundancy is about 5% to about 10%.

10. The method of claim 1, wherein applying the inner codec comprises:

(a) combining symbols from a lane, a symbol history, and a symbol position; and

(b) generating a base candidate using a lookup table, a hash, or both.

11. The method of claim 10, further comprising performing a base repetition check, updating the symbol history, incrementing the lane index, incrementing the frame index, or any combination thereof.

12. The method of claim 1, wherein each lane comprises a plurality of symbols, and applying the inner codec comprises:

(a) generating a base candidate for each symbol within a lane using a lookup table; and

(b) selecting a next lookup table based at least in part on the previously encoded symbol.

13. A method for encoding data, the method comprising:

(a) generating an inner codec comprising a codebook, wherein the codebook is optimized for one or more constraints; and

(b) applying the inner codec to encode data as a plurality of polynucleotide sequences.

14. The method of claim 13, wherein the one or more constraints are related to nucleic acid synthesis, post-processing, storage, sequencing, or any combination thereof.

15. The method of claim 13, further comprising synthesizing a plurality of polynucleotides comprising the plurality of polynucleotide sequences.

16. The method of claim 13, wherein the codebook comprises codewords that are generated based at least in part on a base order.

17. The method of claim 16, wherein the base order comprises predetermined base transitions.

18. The method of claim 13, wherein the inner codec comprises two or more codebooks, and

wherein each of the two or more codebooks encodes a layer during synthesis of the plurality of polynucleotides.

19. The method of claim 18, wherein the layer comprises extension of each polynucleotide of the plurality of polynucleotides by at least one base.

20. The method of claim 15, wherein (c) comprises synthesizing the plurality of polynucleotides on a solid support.