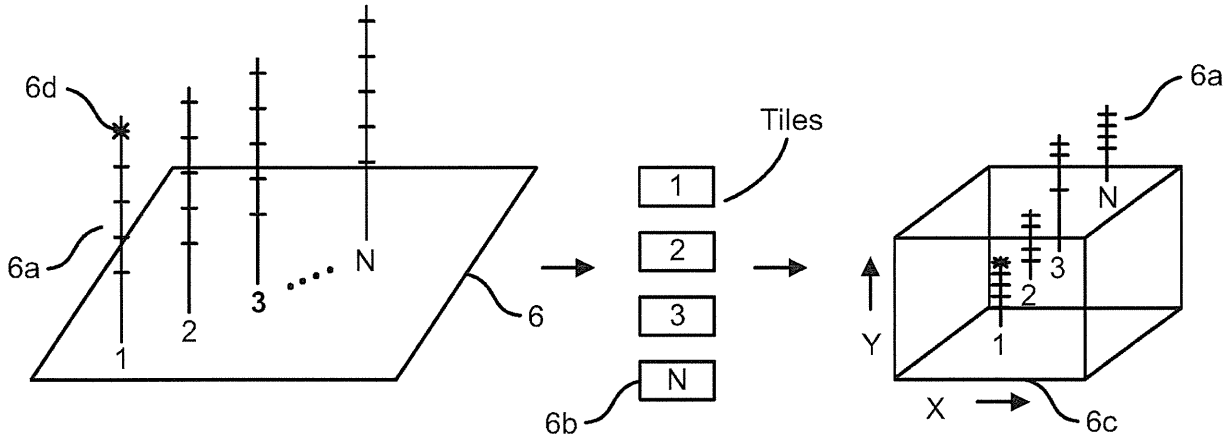


FIG. 1A

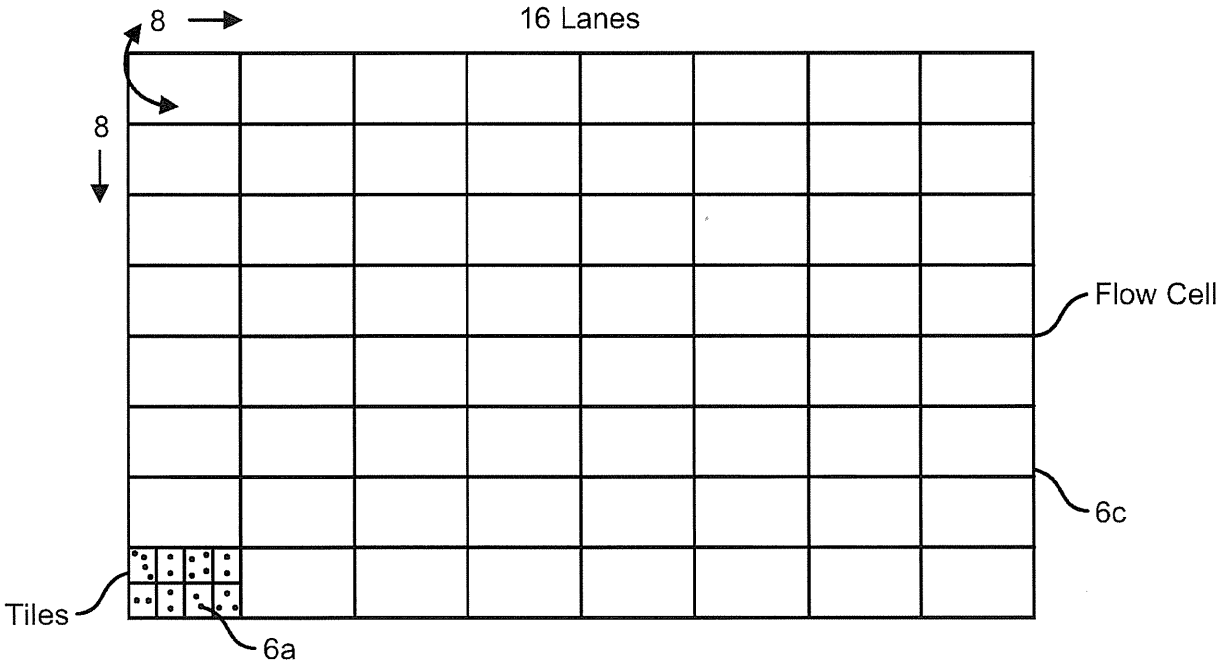


FIG. 1B

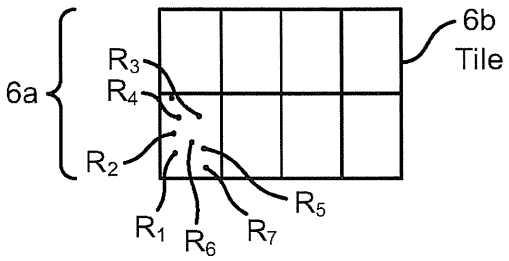


FIG. 1C

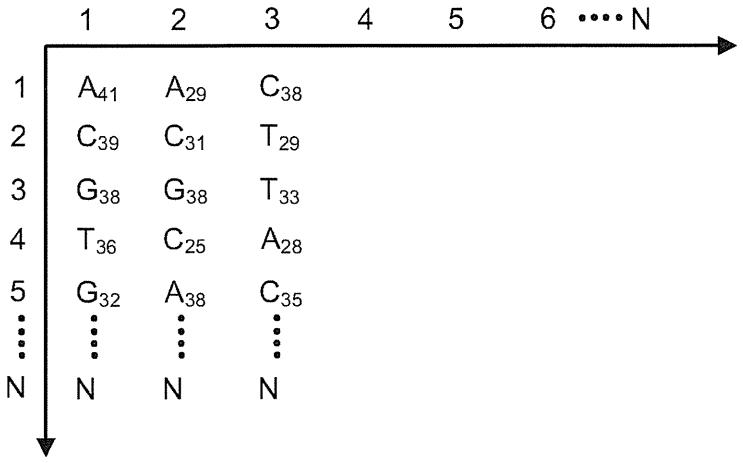


FIG. 1D

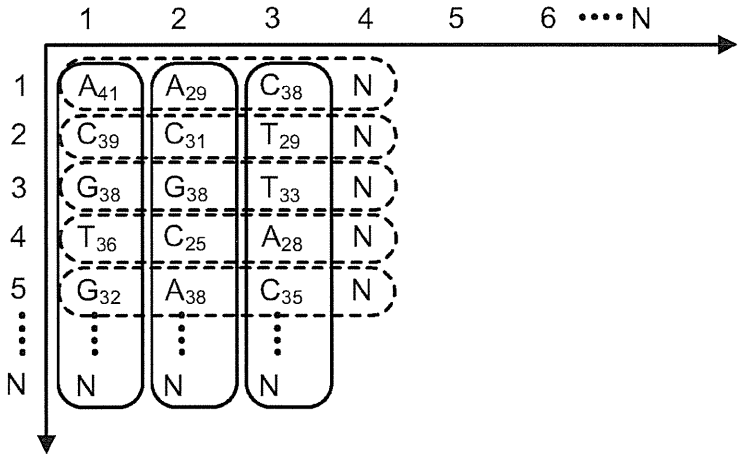


FIG. 1E

Read 1	A_{41}	A_{29}	C_{38}	N
Read 2	C_{39}	C_{31}	T_{29}	N
Read 3	G_{38}	G_{38}	T_{33}	N
Read 4	T_{36}	C_{25}	A_{28}	N
Read 5	G_{32}	A_{38}	C_{35}	N

FIG. 1F

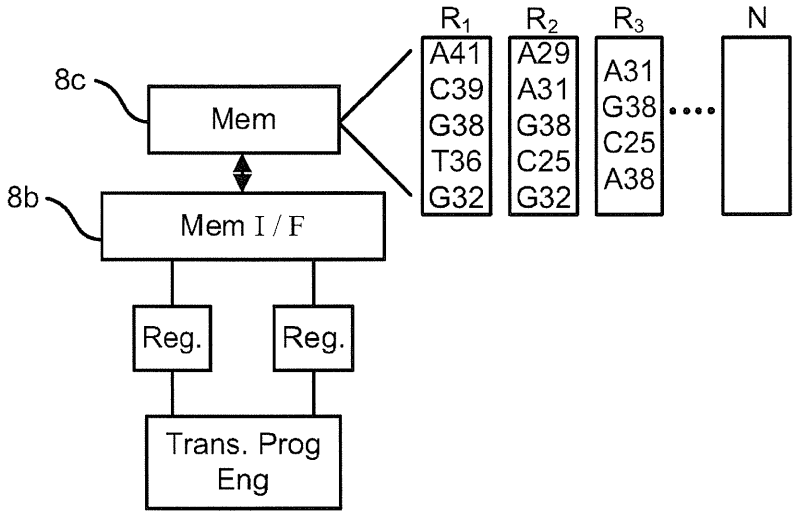


FIG. 1G

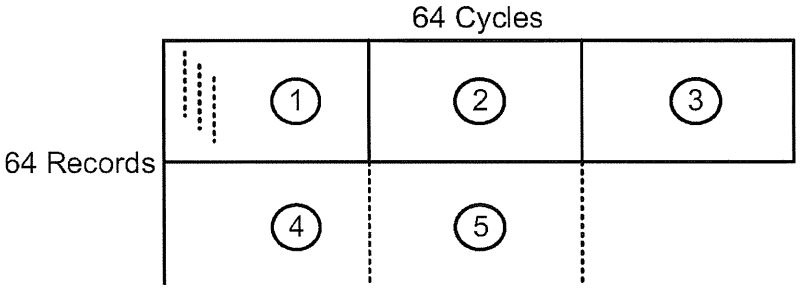


FIG. 1H

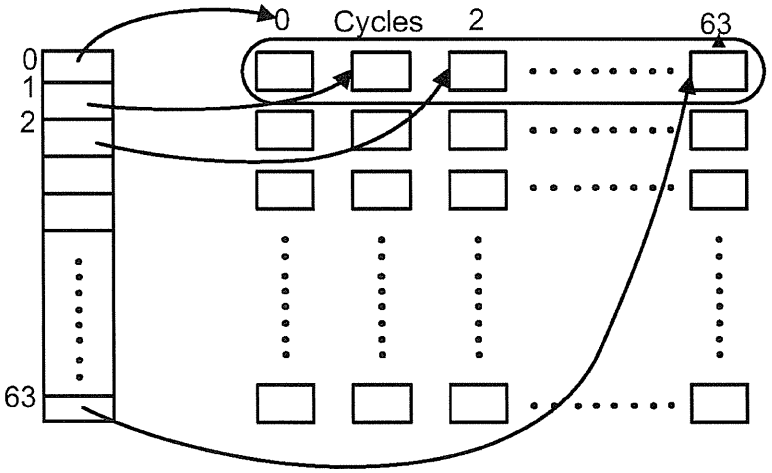


FIG. 1I

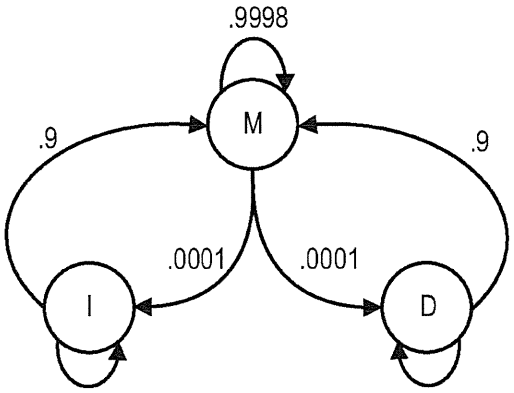


FIG. 2

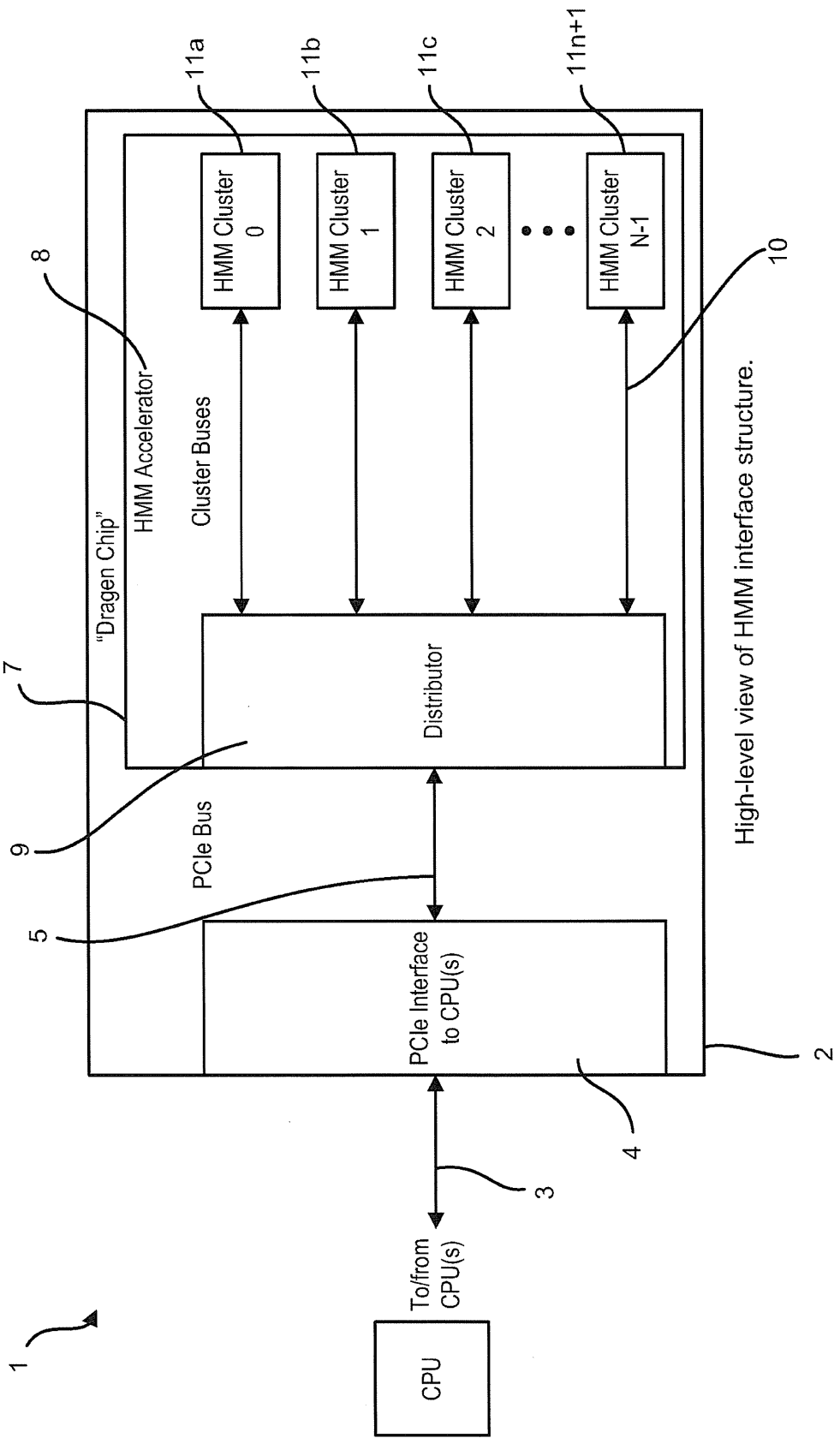

1
CPU
To/from CPU(s)
3
4
2
PCIe Interface to CPU(s)
5
PCIe Bus
9
7
"Dragen Chip"
HMM Accelerator
8
Cluster Buses
HMM Cluster 0
HMM Cluster 1
HMM Cluster 2
HMM Cluster N-1
11a
11b
11c
11n+1
Distributor
10
High-level view of HMM interface structure.

FIG. 3A

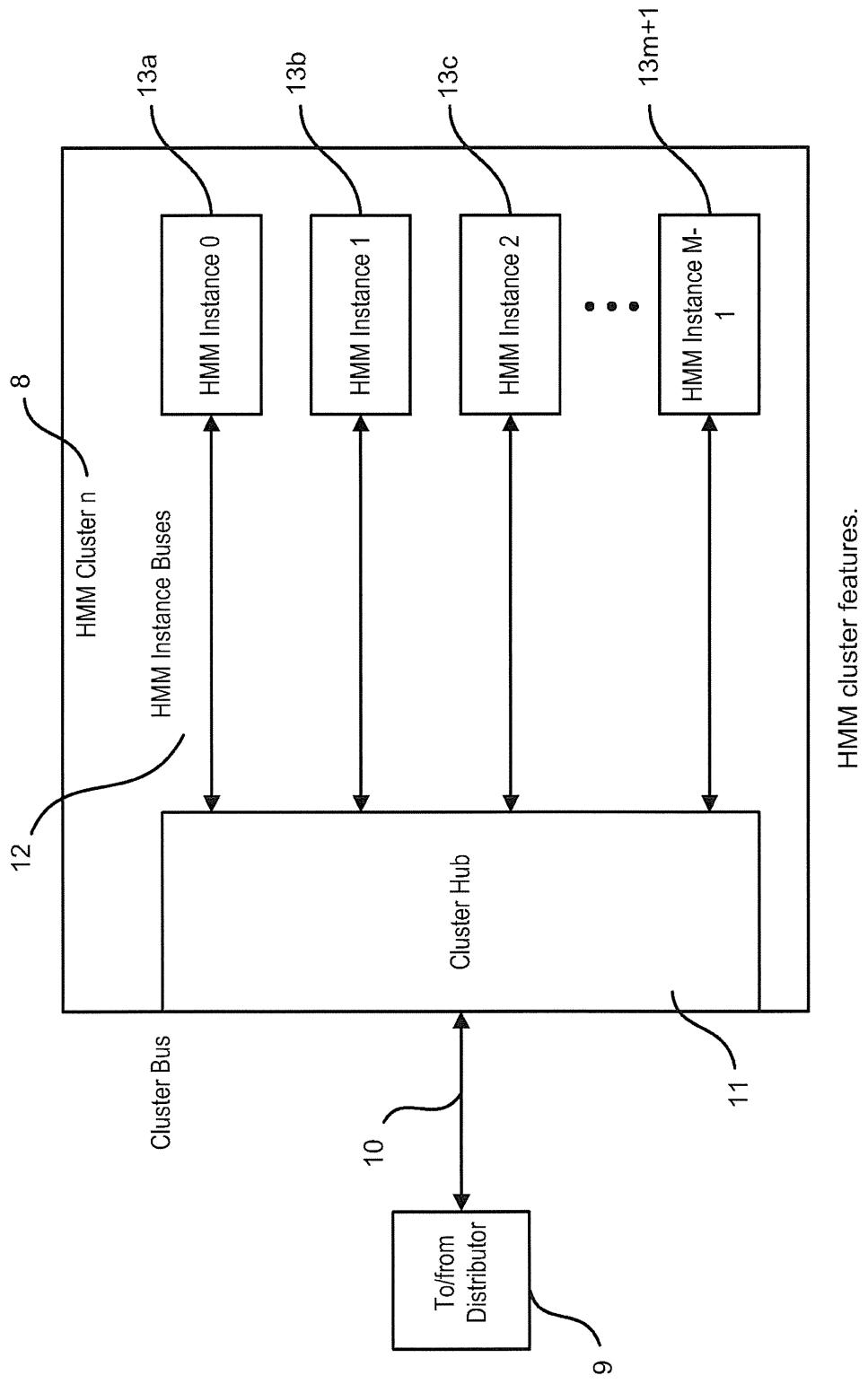

HMM Cluster n
8
HMM Instance Buses
12
Cluster Hub
HMM Instance 0
HMM Instance 1
HMM Instance 2
HMM Instance M-1
13a
13b
13c
13m+1
HMM cluster features.
Cluster Bus
10
11
To/from Distributor
9

FIG. 3B

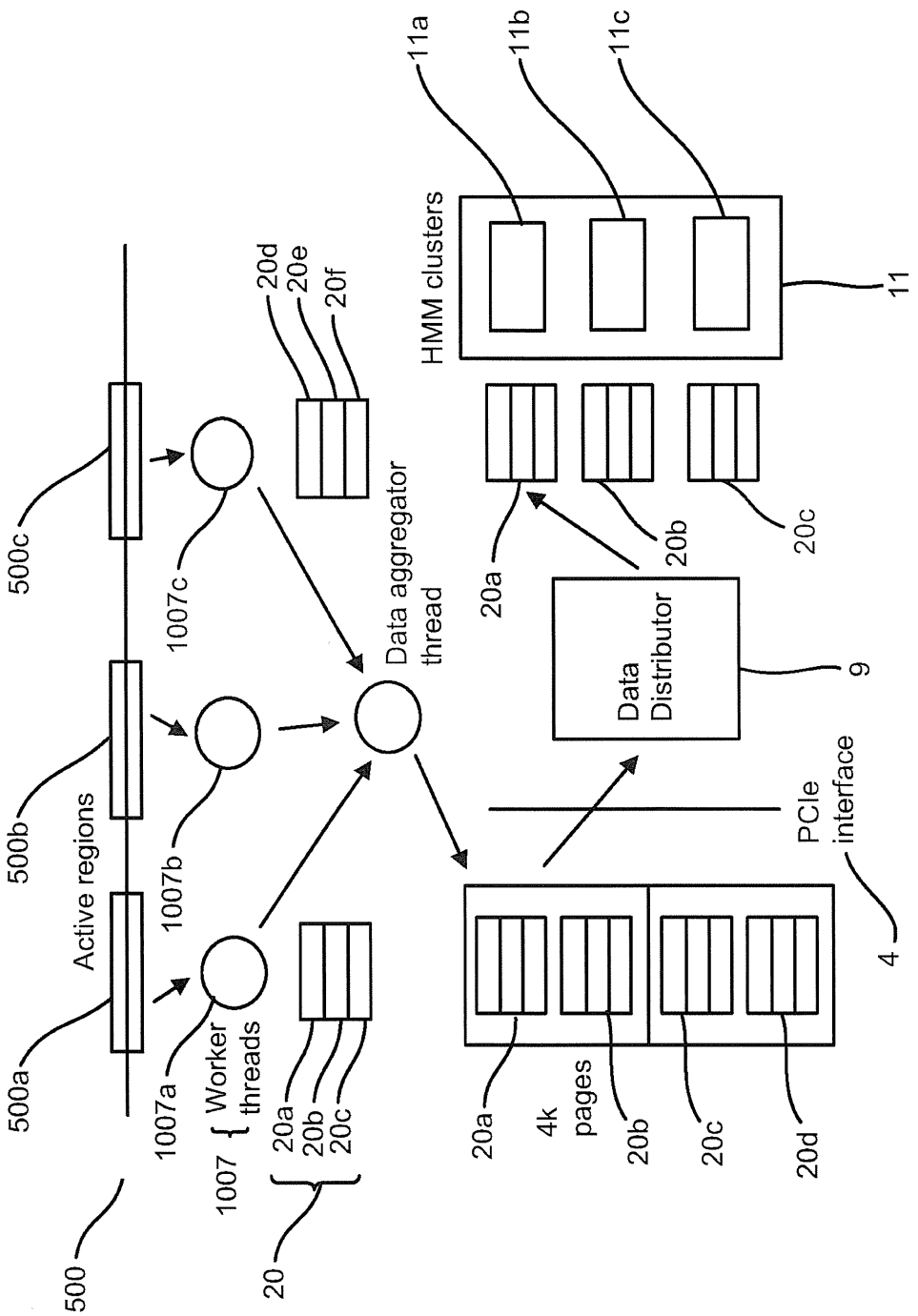
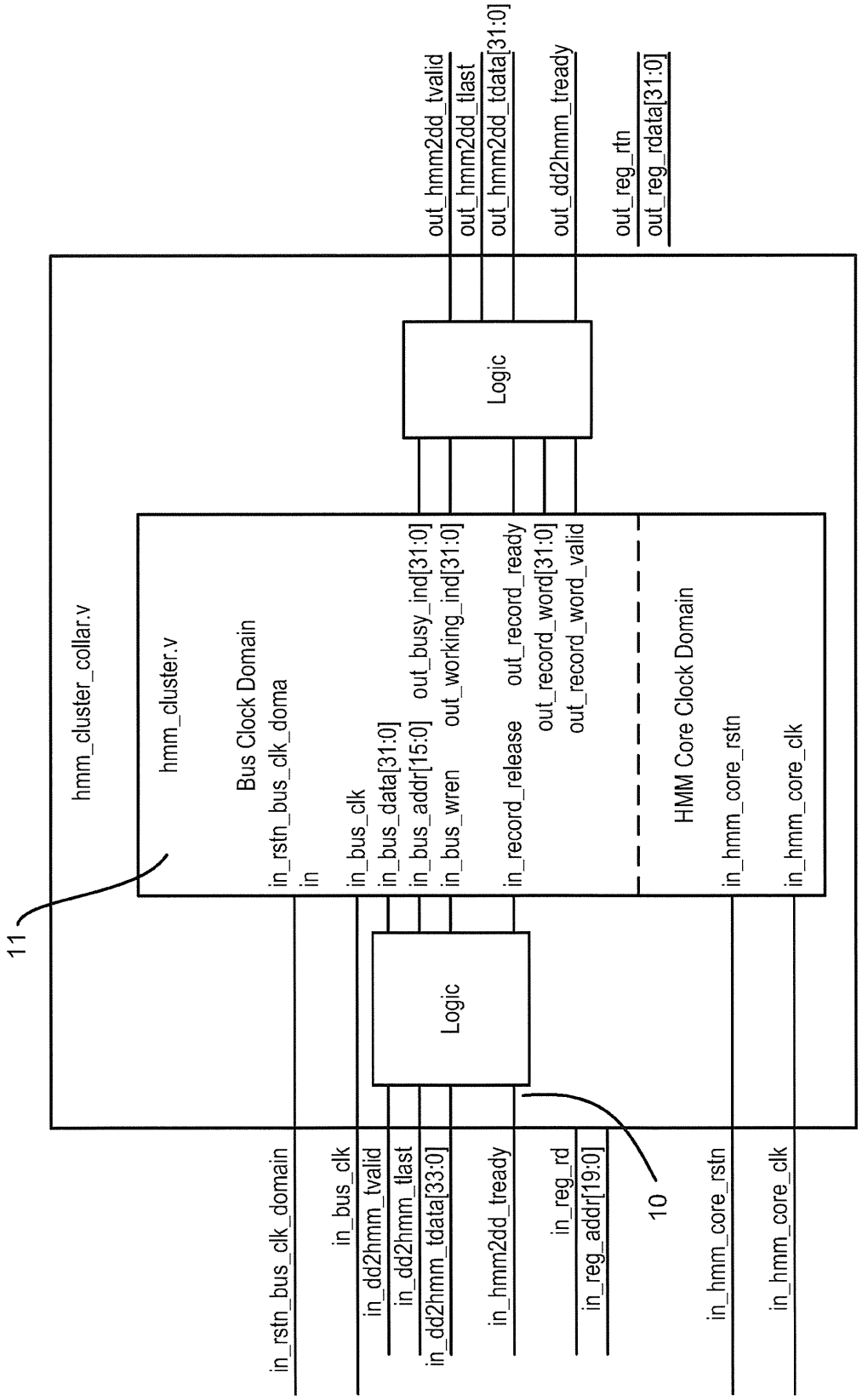


FIG. 4

HMM data flow and HW/SW interaction overview.



HMM Cluster Collar connections.

FIG. 5

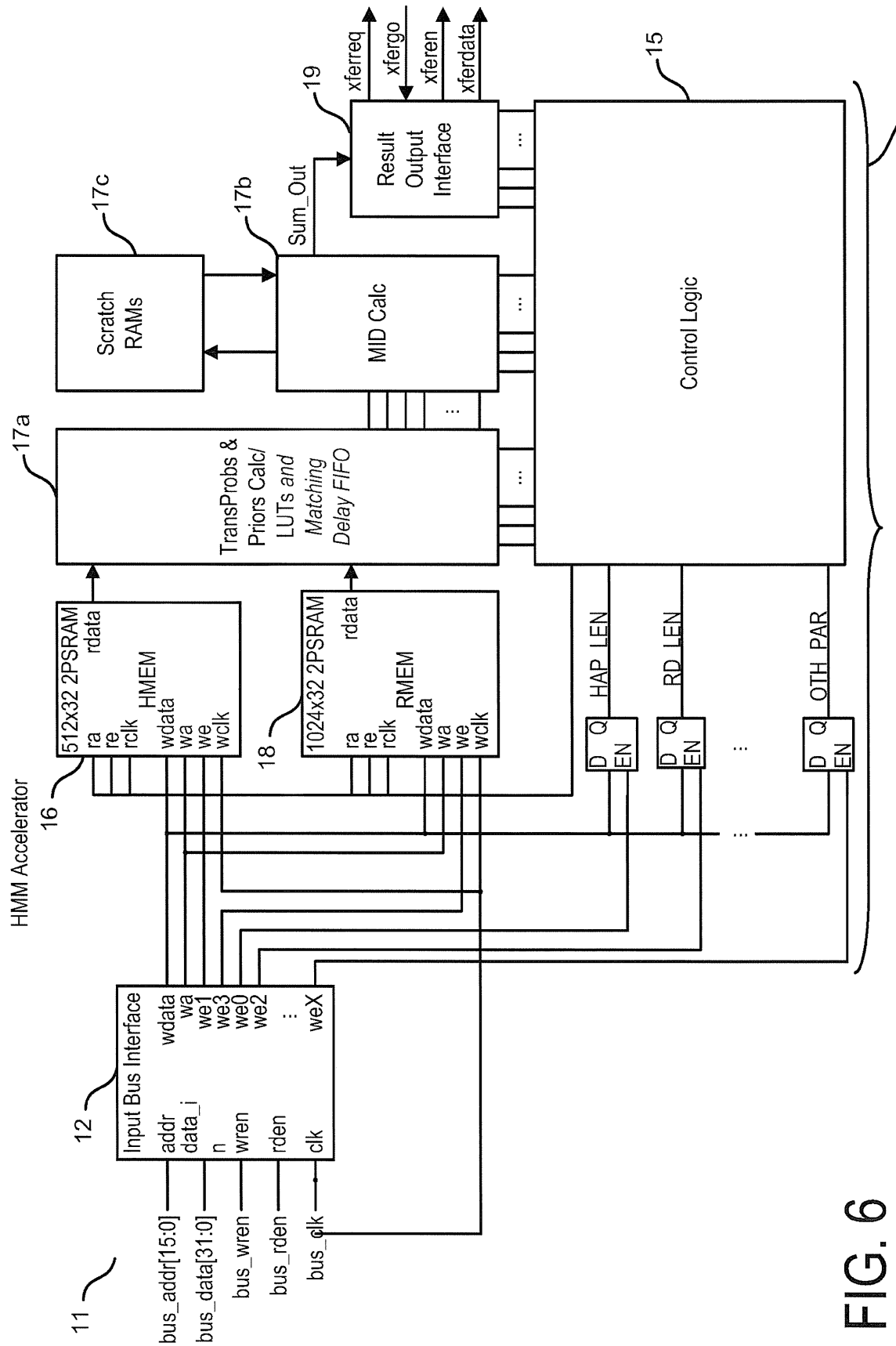
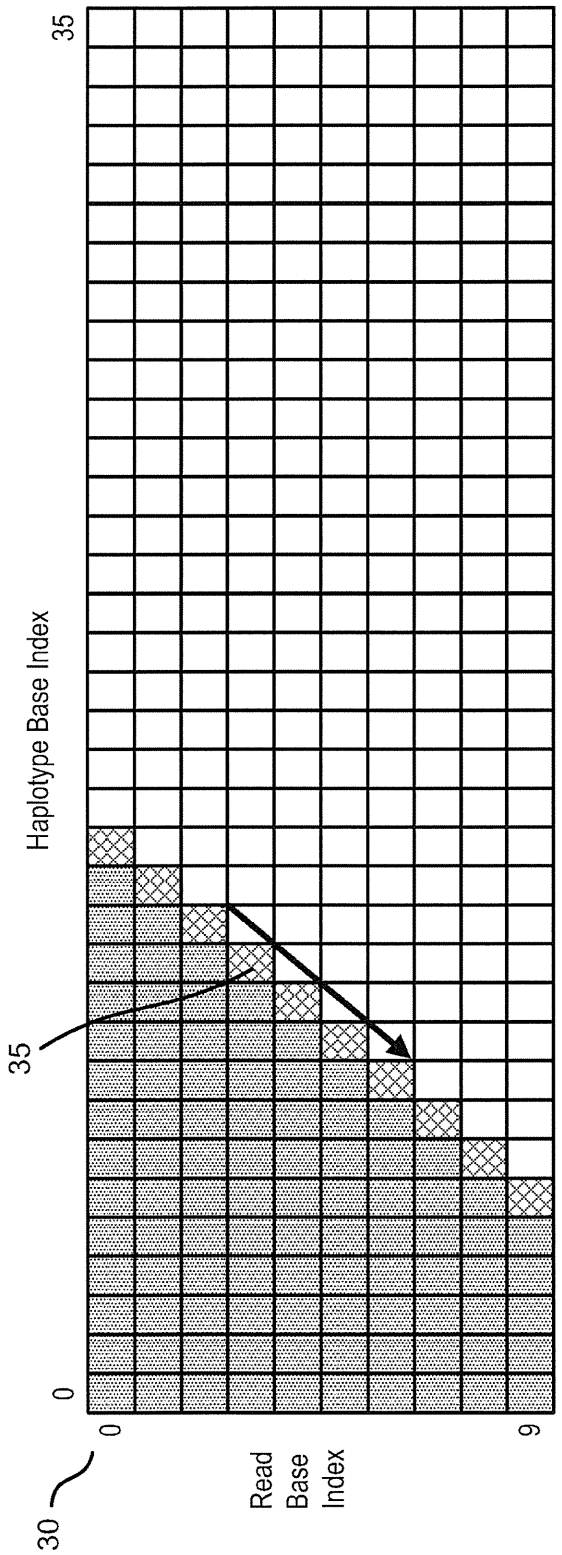


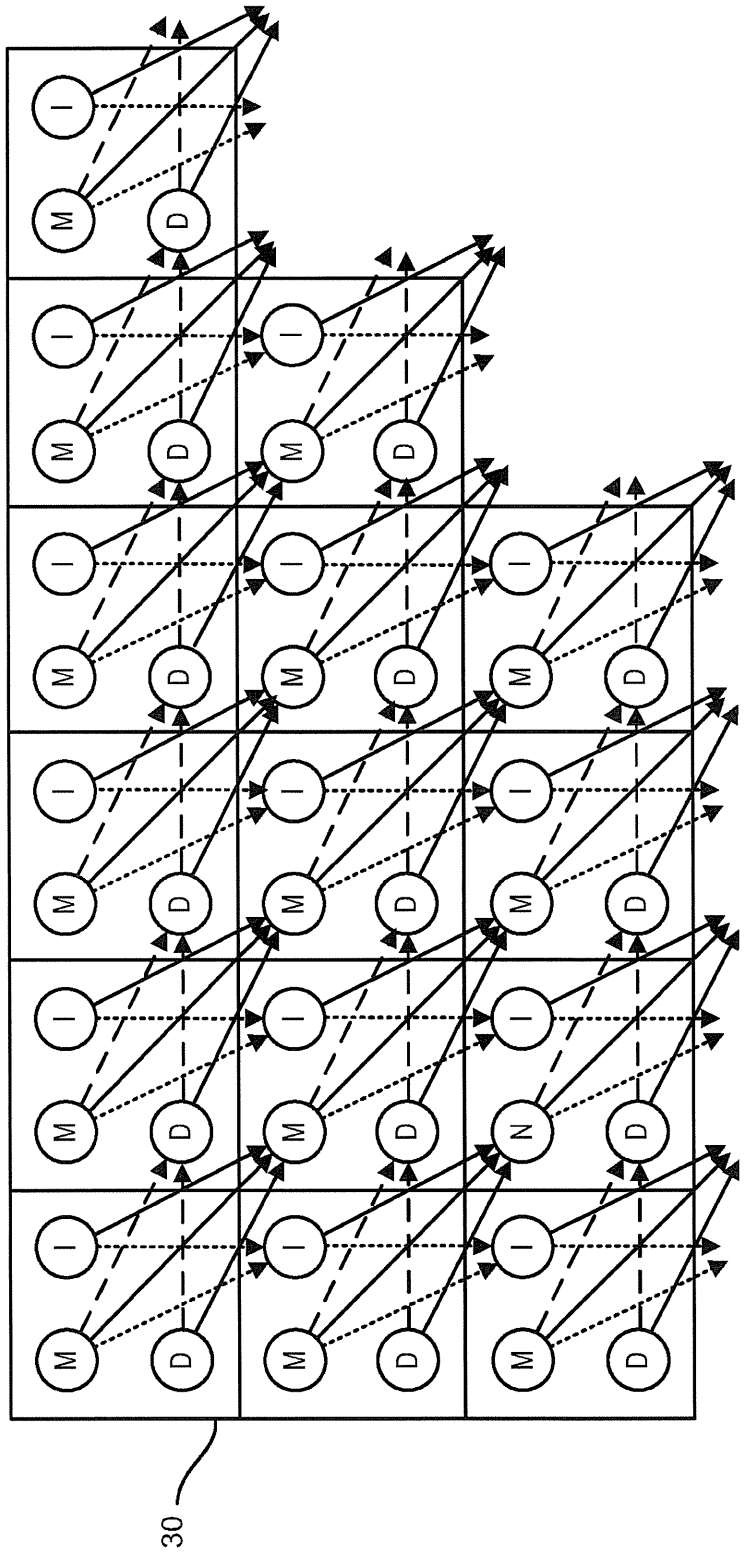
FIG. 6

Somewhat high-level view of the major functional blocks within the HMM hardware accelerator.



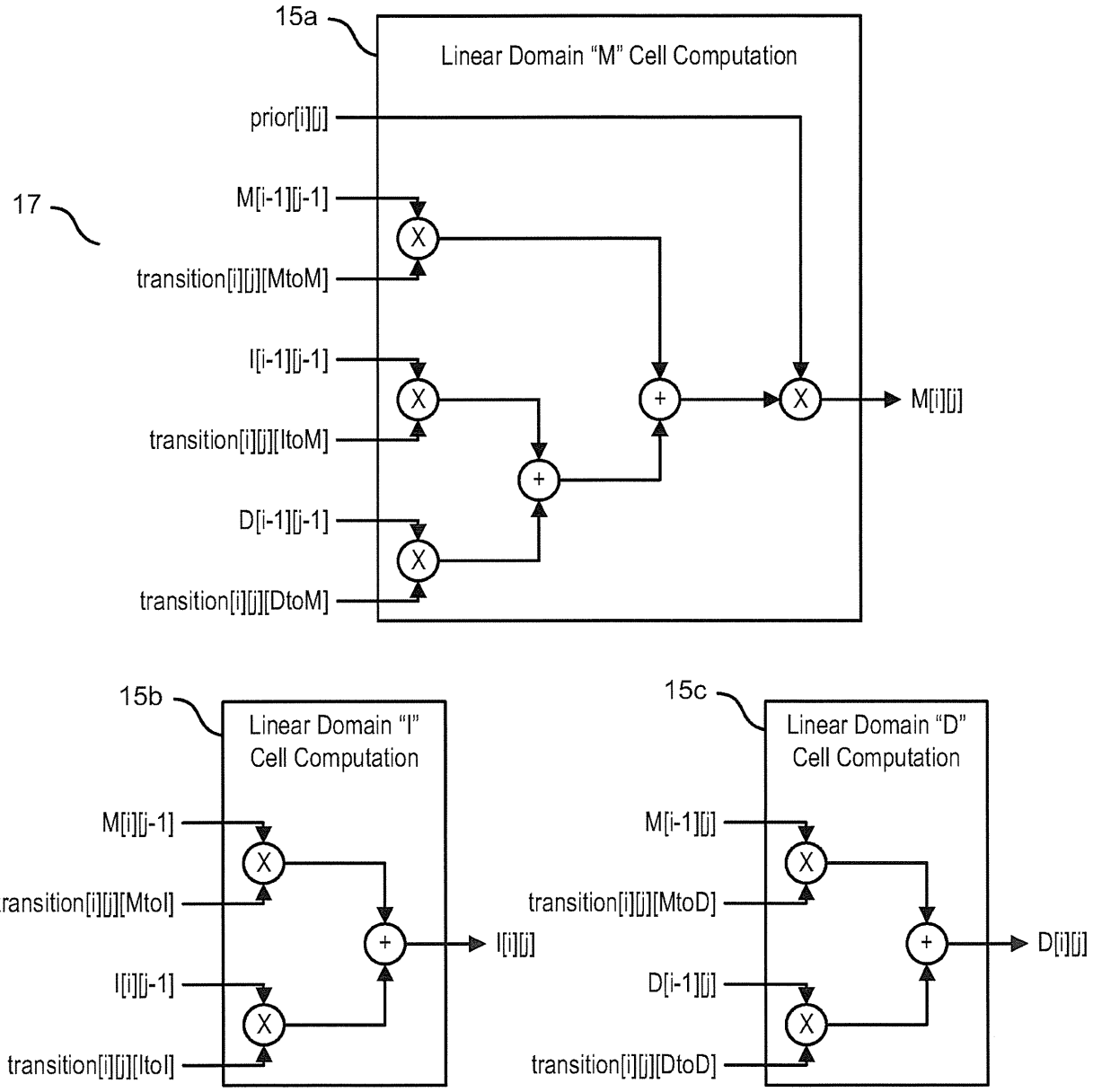
HMM matrix structure and hardware processing flow example.

FIG. 7



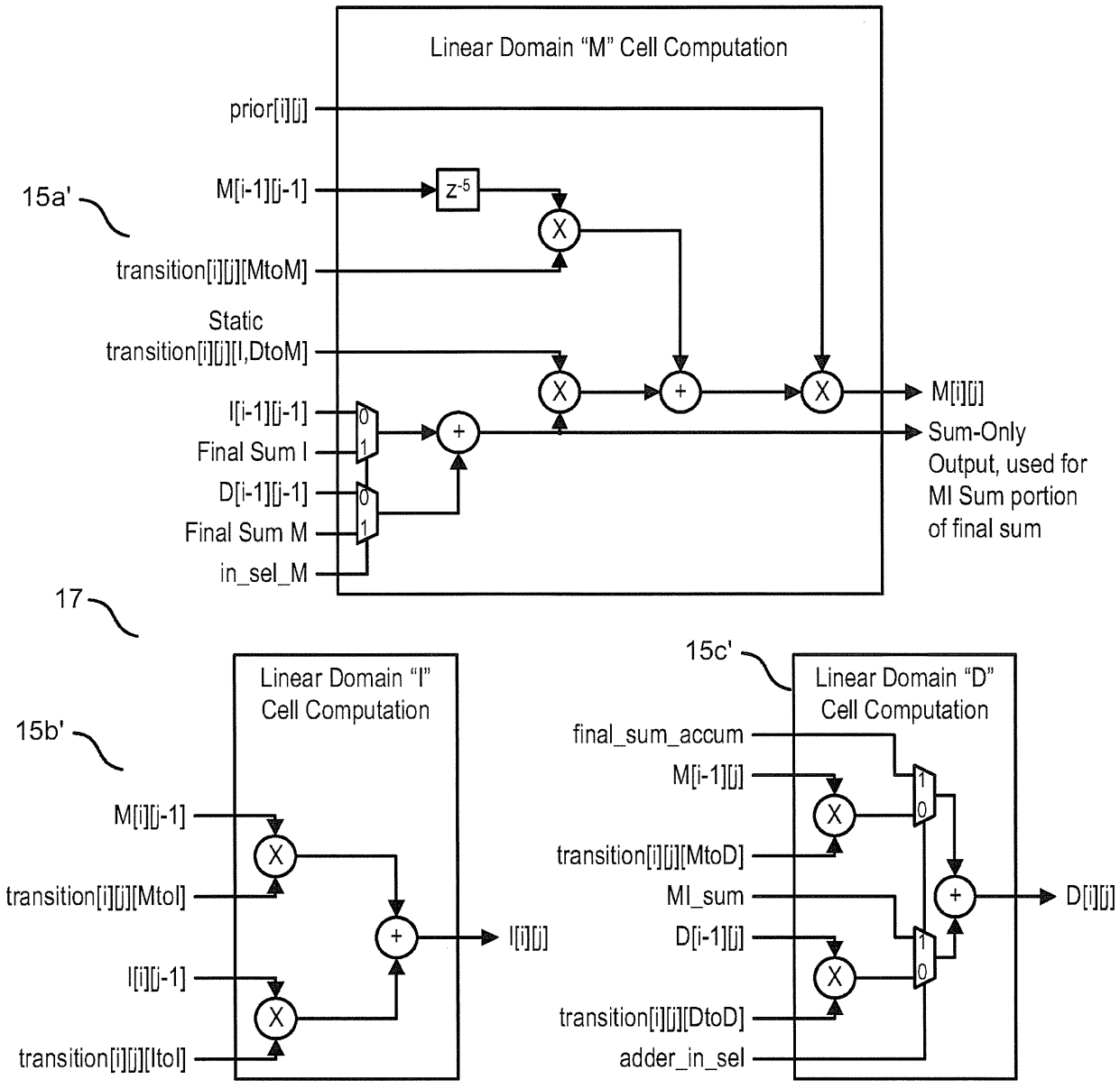
Enlarged view of a portion of Figure 8 showing the data flow and dependencies between nearby cells in the HMM M,I, and D state computations.

FIG. 8



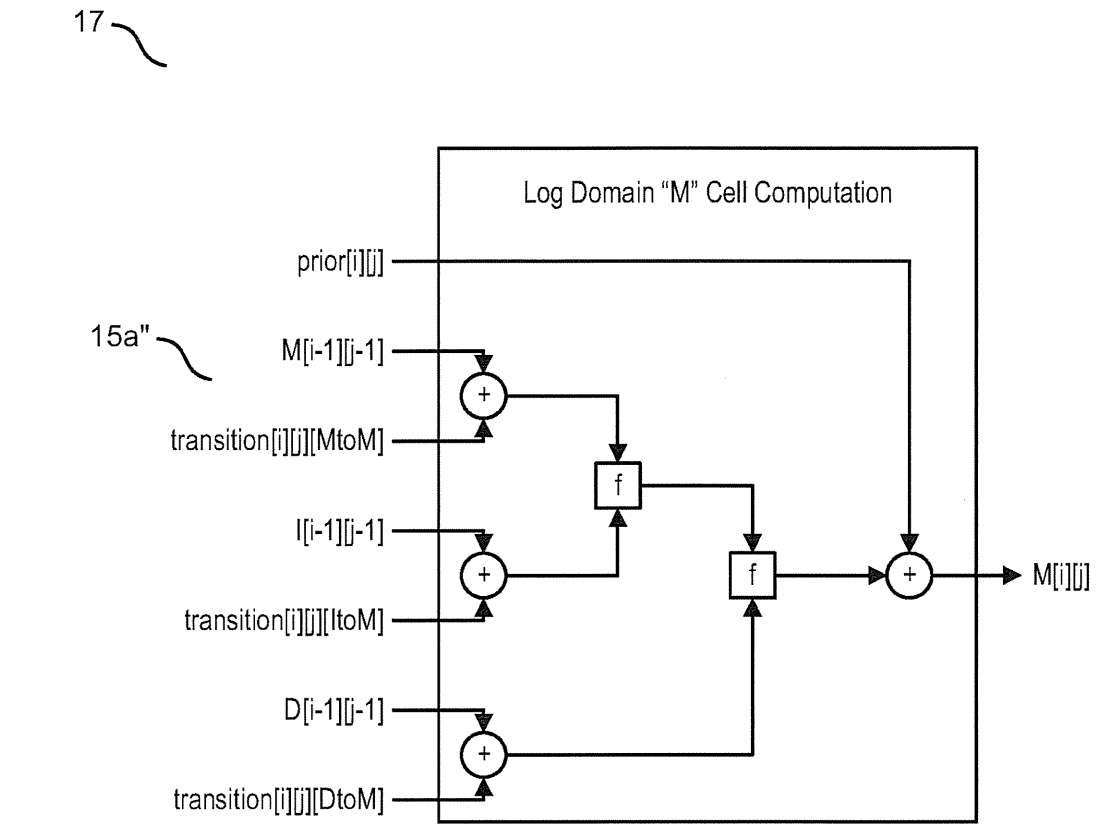
Computations required for M, I, D state updates.

FIG. 9

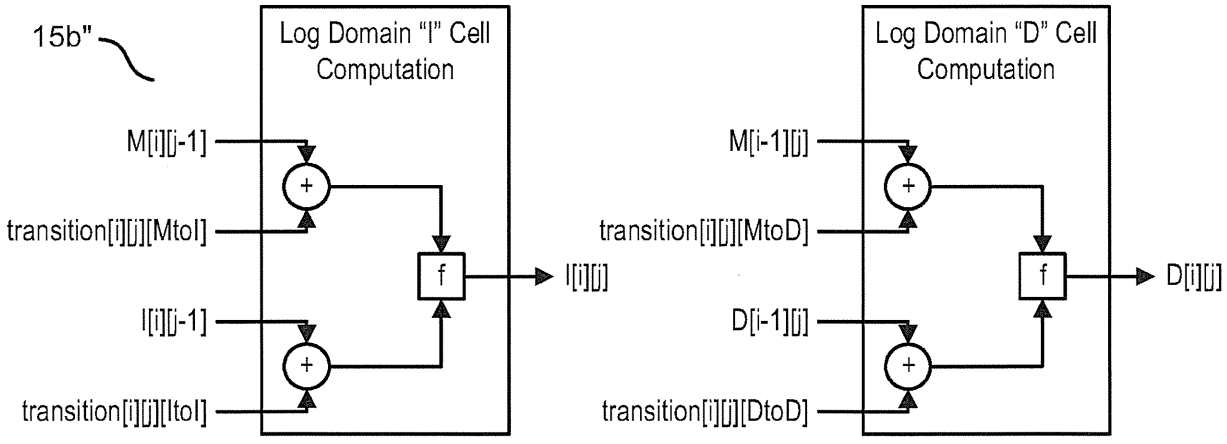


M,1, and D state update circuits, including effects of simplifying assumptions related to transition probabilities and the effect of sharing some M, I, D adder resources with the final sum operations.

FIG. 10

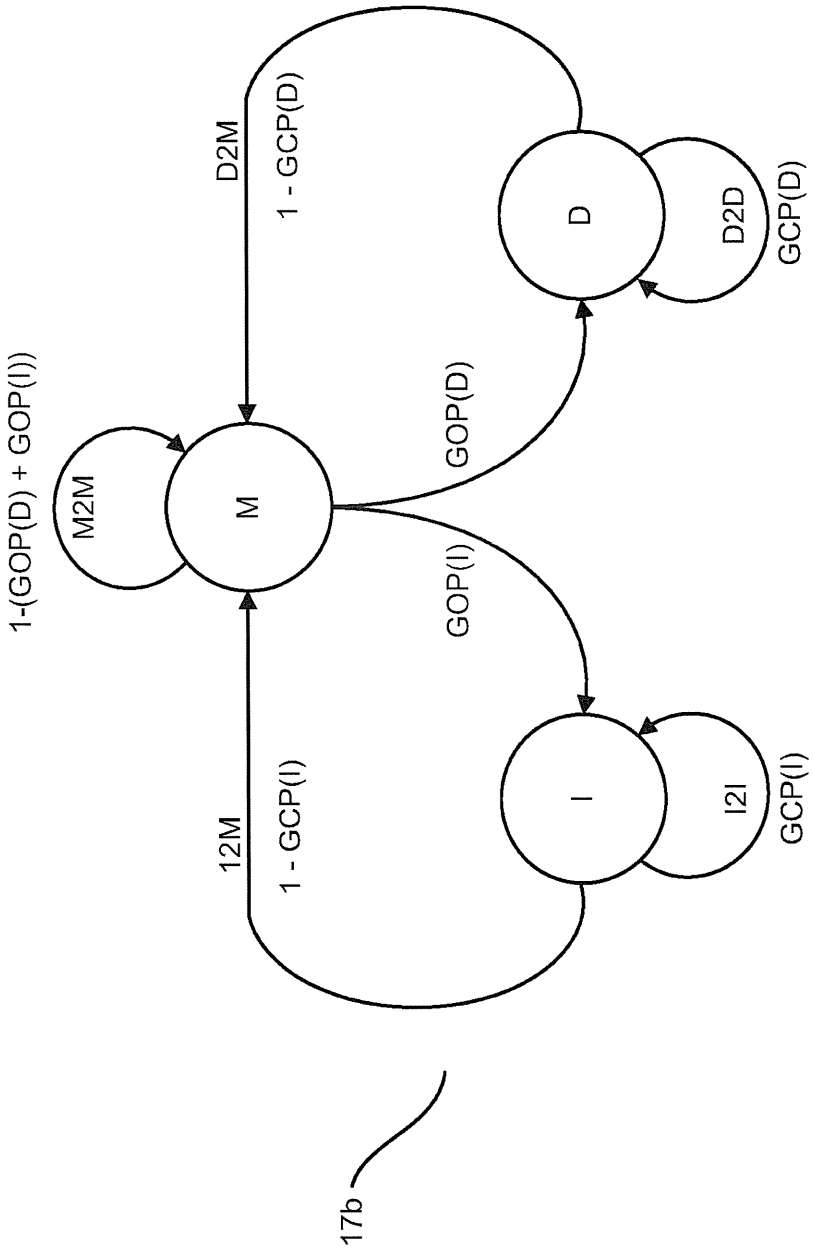


Note: The "f" function is the approximation to log of addition. I.e., $f(a,b) \approx \max(a,b) - \log_2(1+2^{(-|a-b|)})$



Log domain M, I, D state calculation details.

FIG. 11



HMM state transition diagram showing relation between GOP, GCP and transition probabilities.

FIG. 12

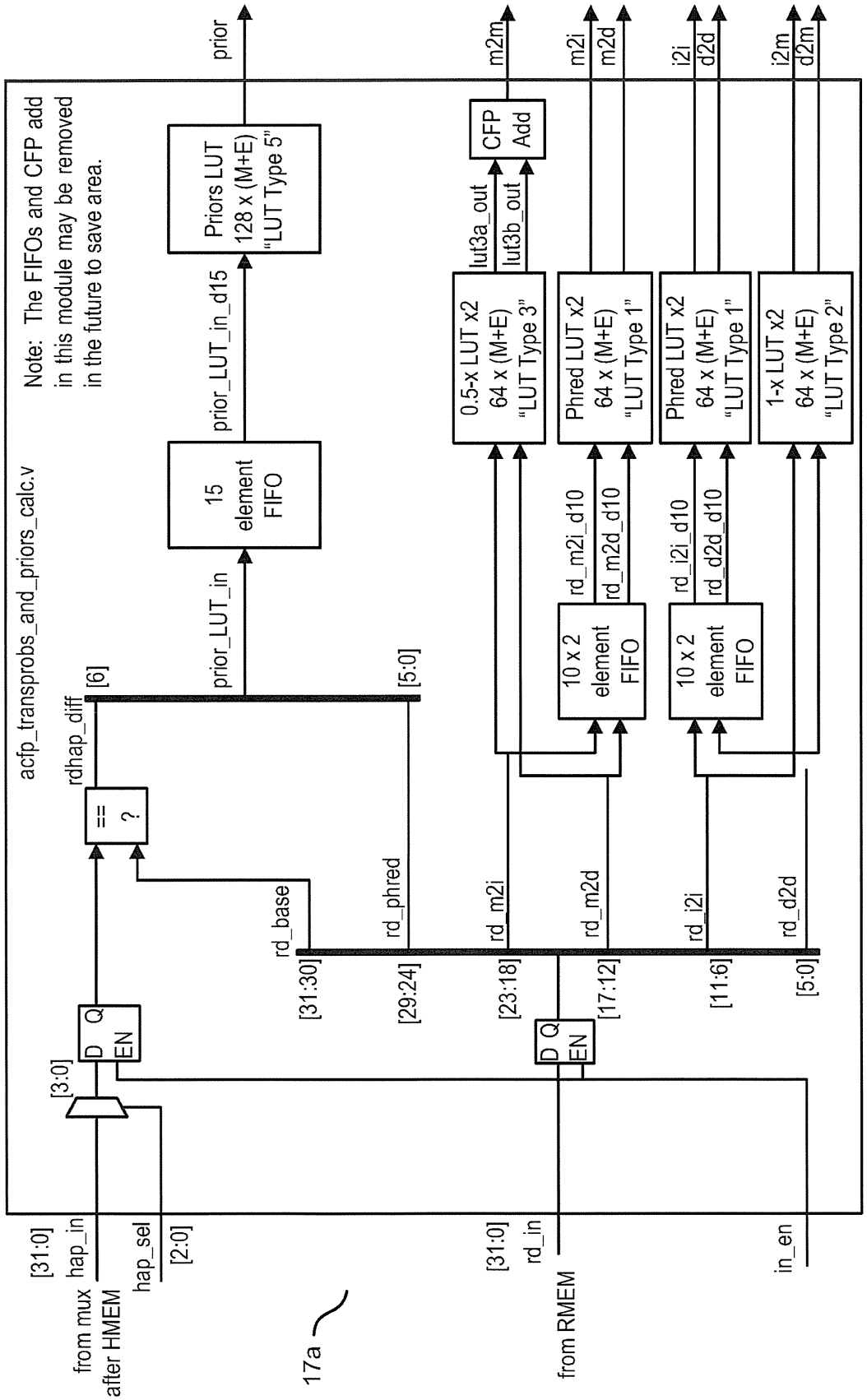
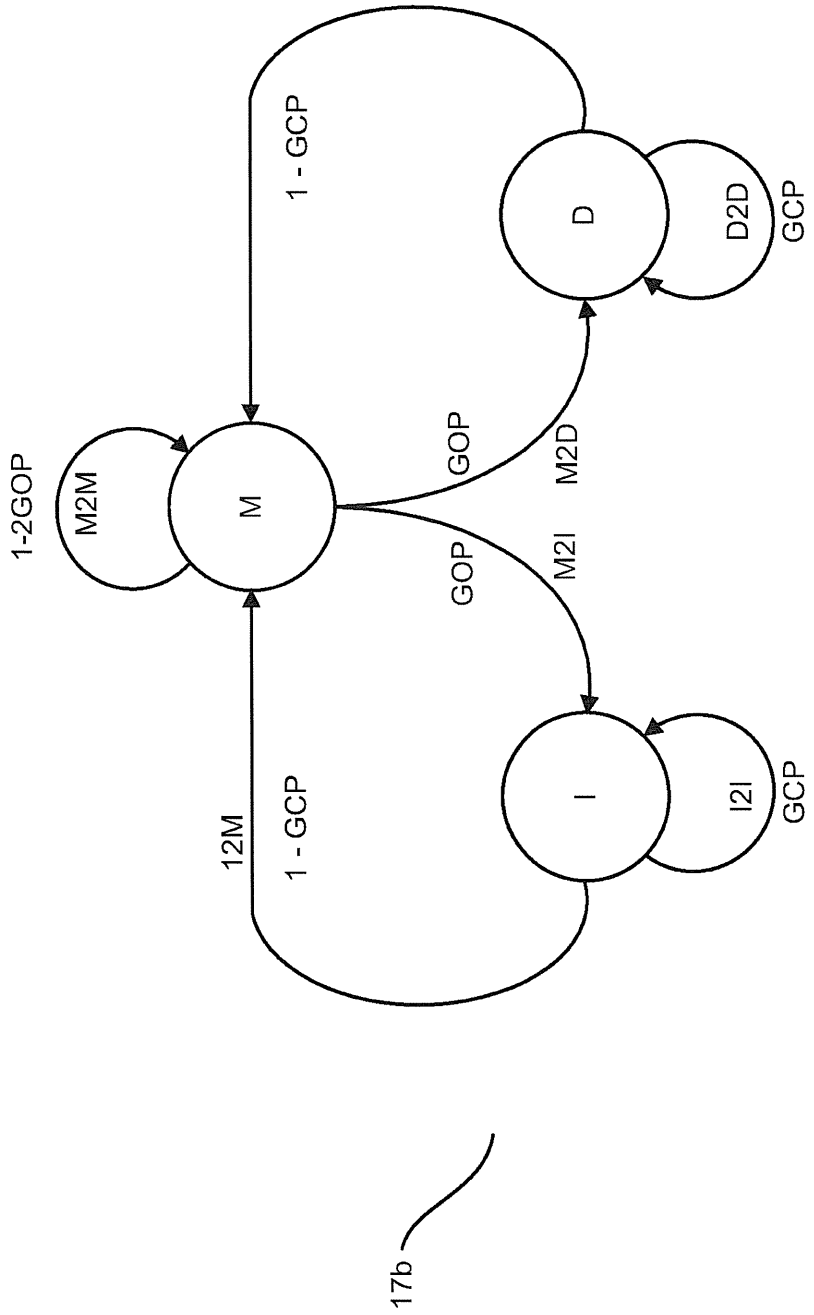


FIG. 13

HMM Transprobs and priors generation circuit to support the general state transaction diagram of Figure 17.



Simplified HMM state transition diagram showing relation between GOP, GCP and transition probabilities.

FIG. 14

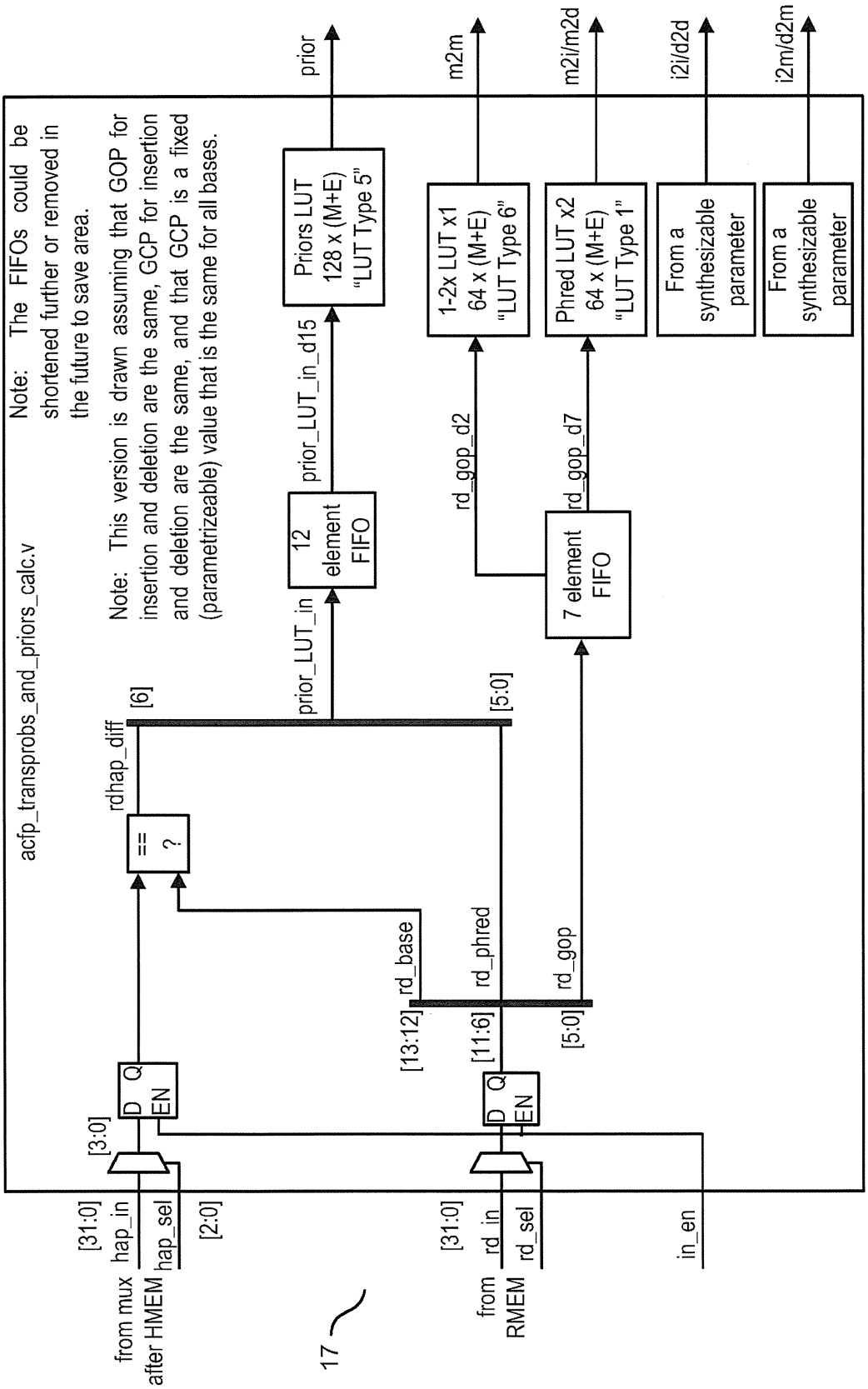


FIG. 15

HMM transprobs and priors generation circuit to support the simplified state transition diagram of Figure 19.

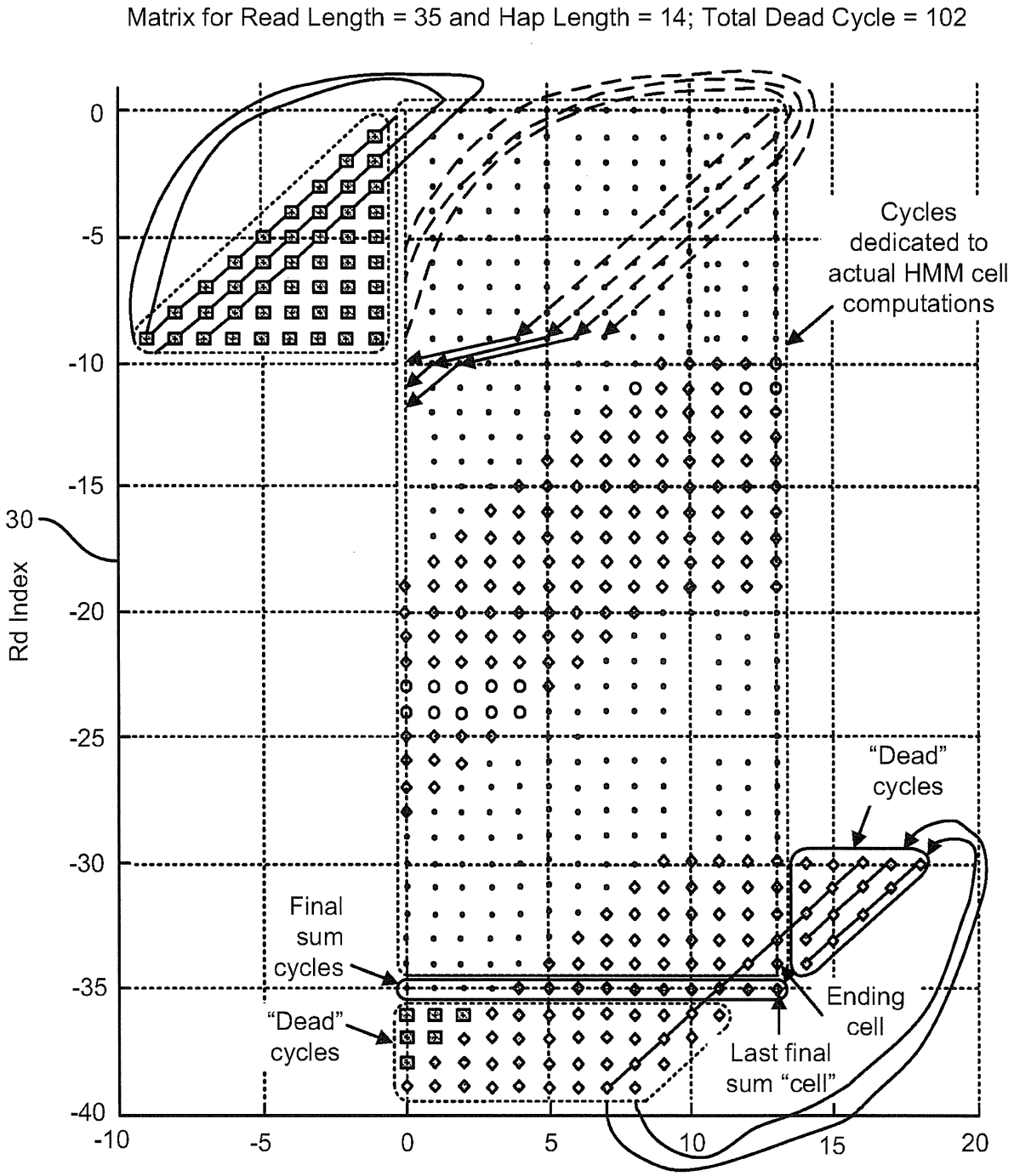

Matrix for Read Length = 35 and Hap Length = 14; Total Dead Cycle = 102
Rd Index
0
-5
-10
-15
30
-20
-25
-30
-35
-40
-10
-5
0
5
10
15
20
Cycles dedicated to actual HMM cell computations
"Dead" cycles
Final sum cycles
"Dead" cycles
Ending cell
Last final sum "cell"

FIG. 16

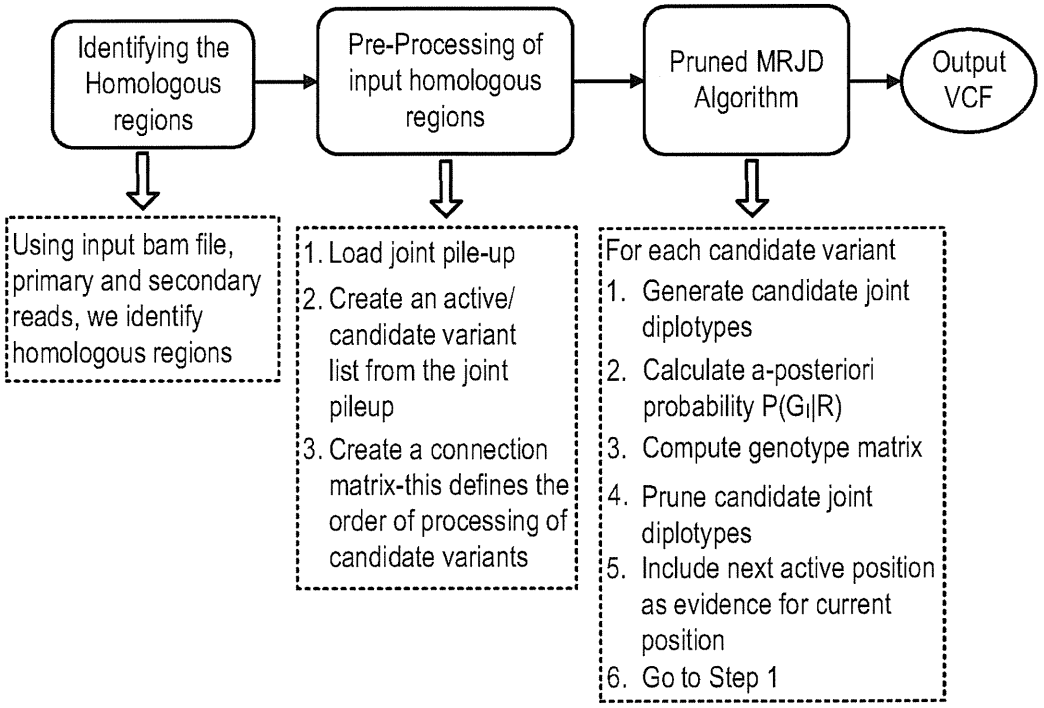


FIG. 17A

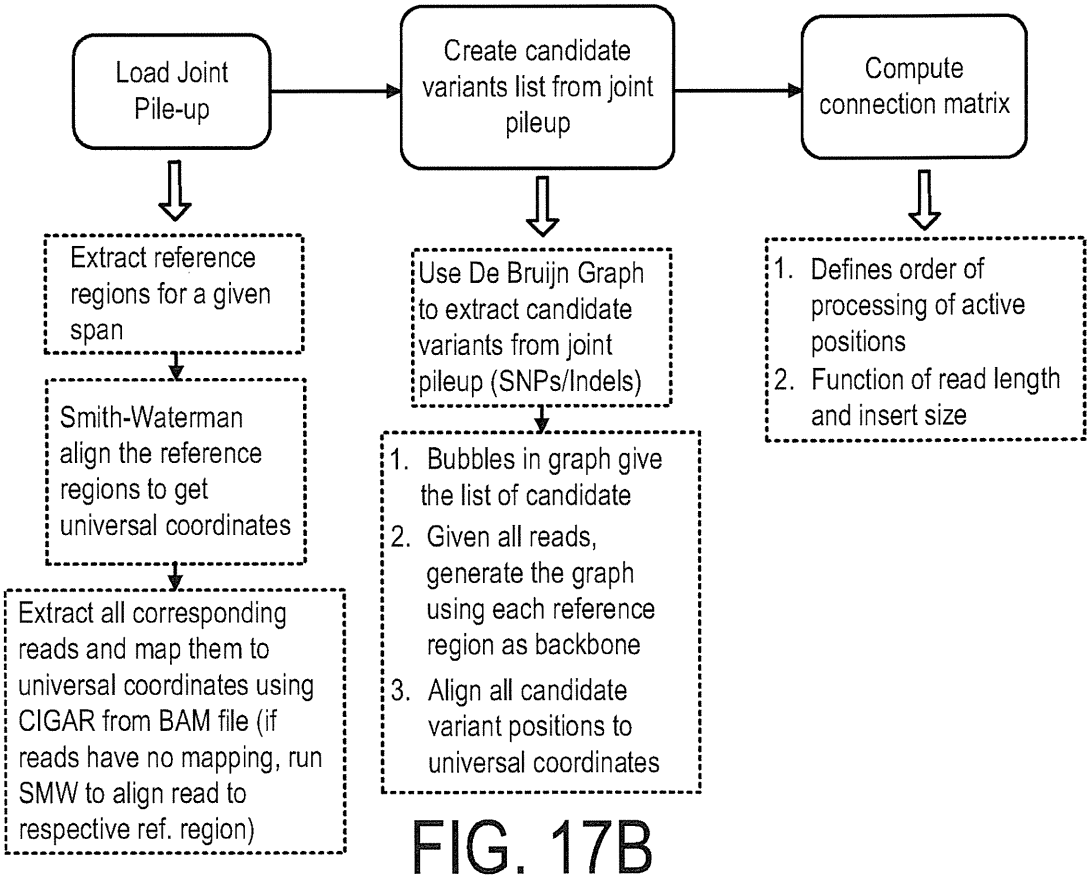


FIG. 17B

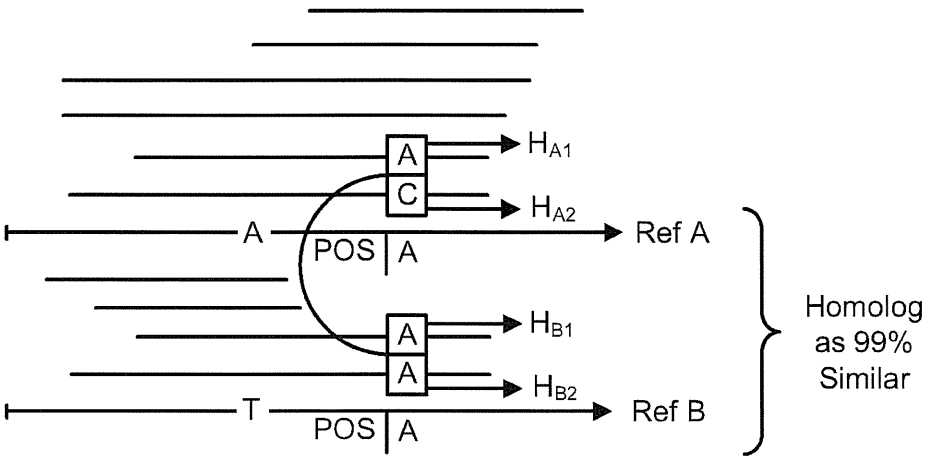


FIG. 18A

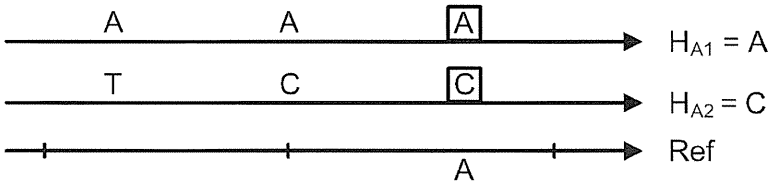


FIG. 18B

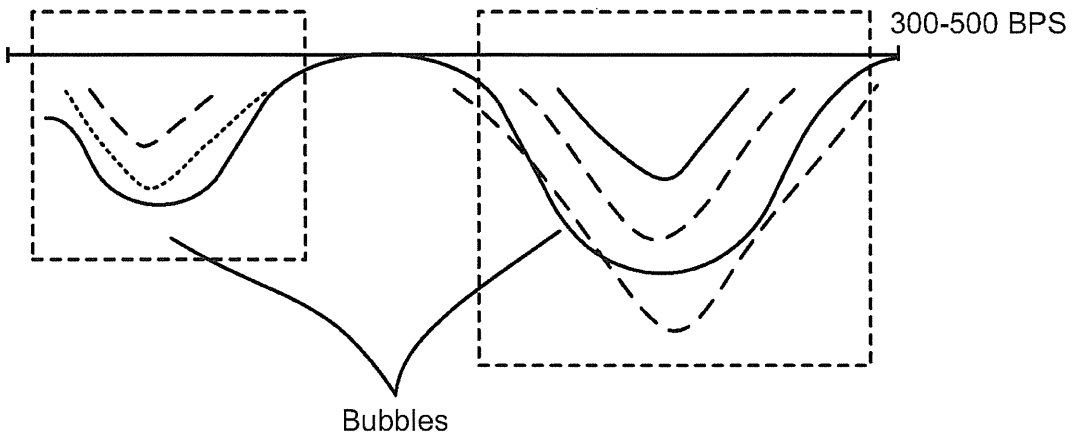


FIG. 18C

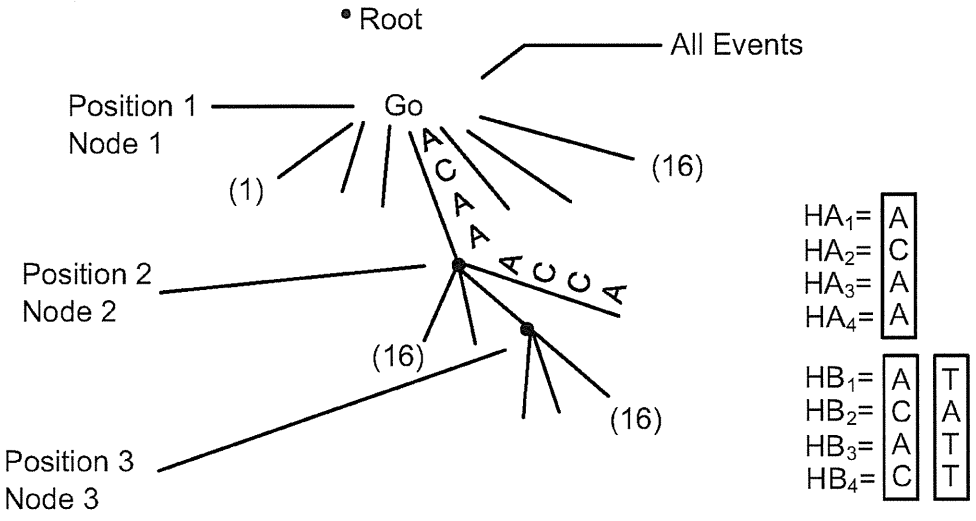


FIG. 18D

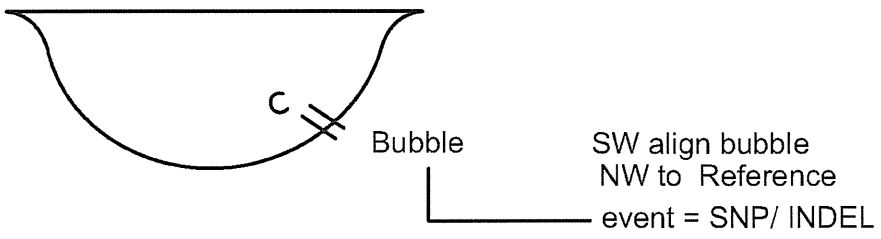


FIG. 18E

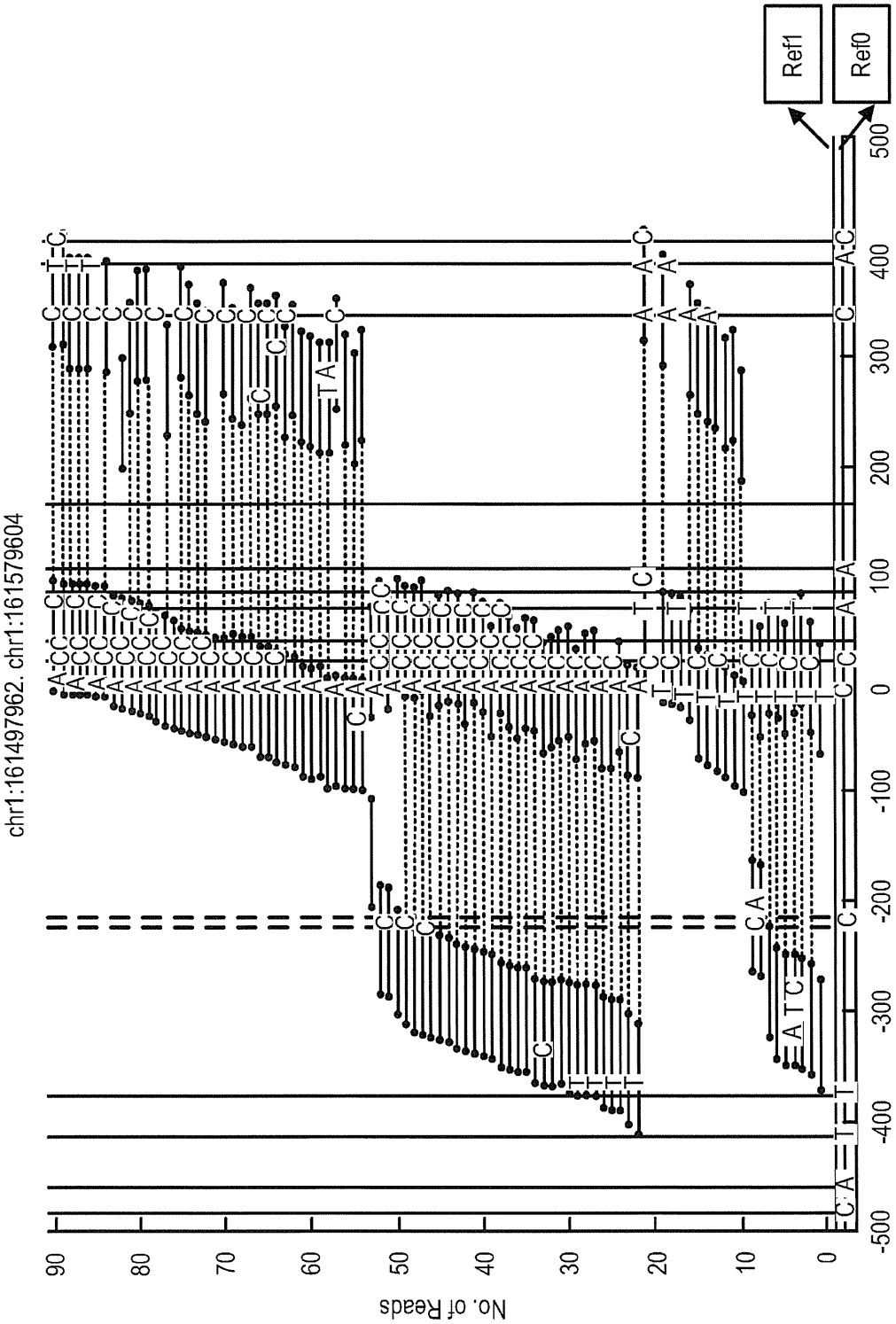

chr1:161497962. chr1:161579604
Ref1
Ref0
No. of Reads
Base Index
90
80
70
60
50
40
30
20
10
0
-500
-400
-300
-200
-100
0
100
200
300
400
500

FIG. 19

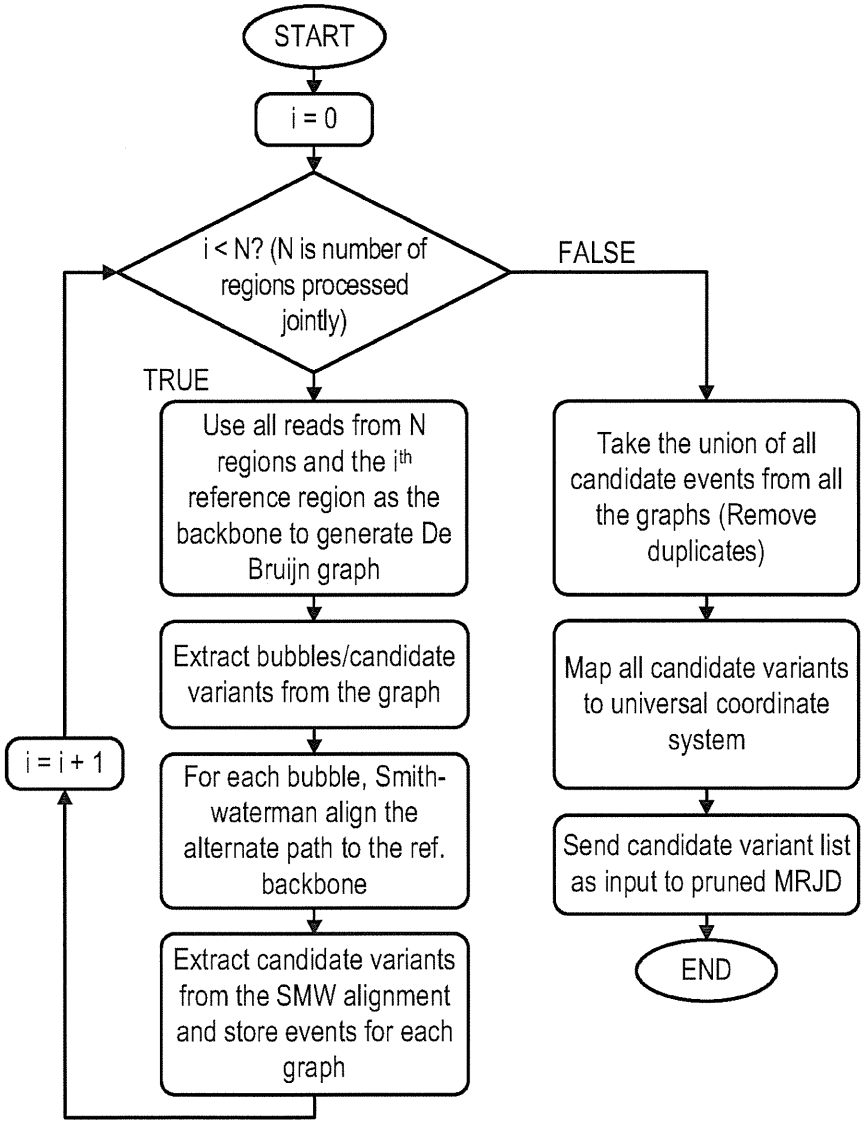

START
i = 0
i < N? (N is number of regions processed jointly)
FALSE
TRUE
Use all reads from N regions and the ith reference region as the backbone to generate De Bruijn graph
Extract bubbles/candidate variants from the graph
For each bubble, Smith-waterman align the alternate path to the ref. backbone
Extract candidate variants from the SMW alignment and store events for each graph
i = i + 1
Take the union of all candidate events from all the graphs (Remove duplicates)
Map all candidate variants to universal coordinate system
Send candidate variant list as input to pruned MRJD
END

FIG. 20

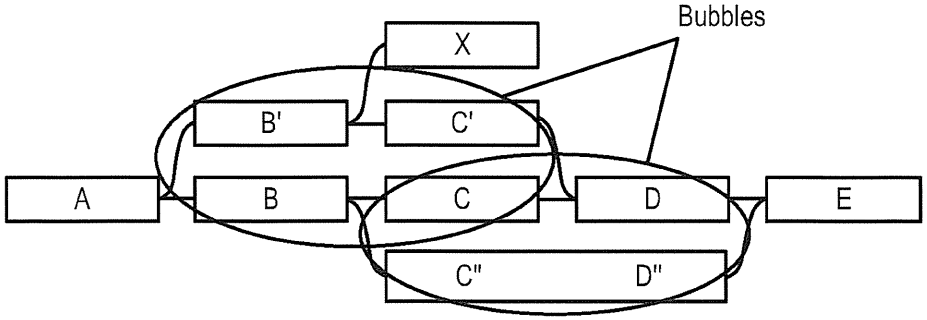
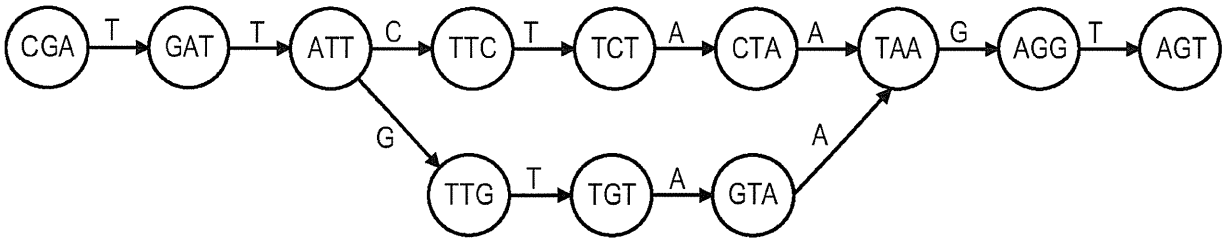

Bubbles
X
B'
C'
A
B
C
D
E
C"
D"

FIG. 21



Candidate haplotypes	Bubble Extraction
CGATT**C**TAAGT	TT**C**TAA
CGATT**G**TAAGT	TT**G**TAA

FIG. 22

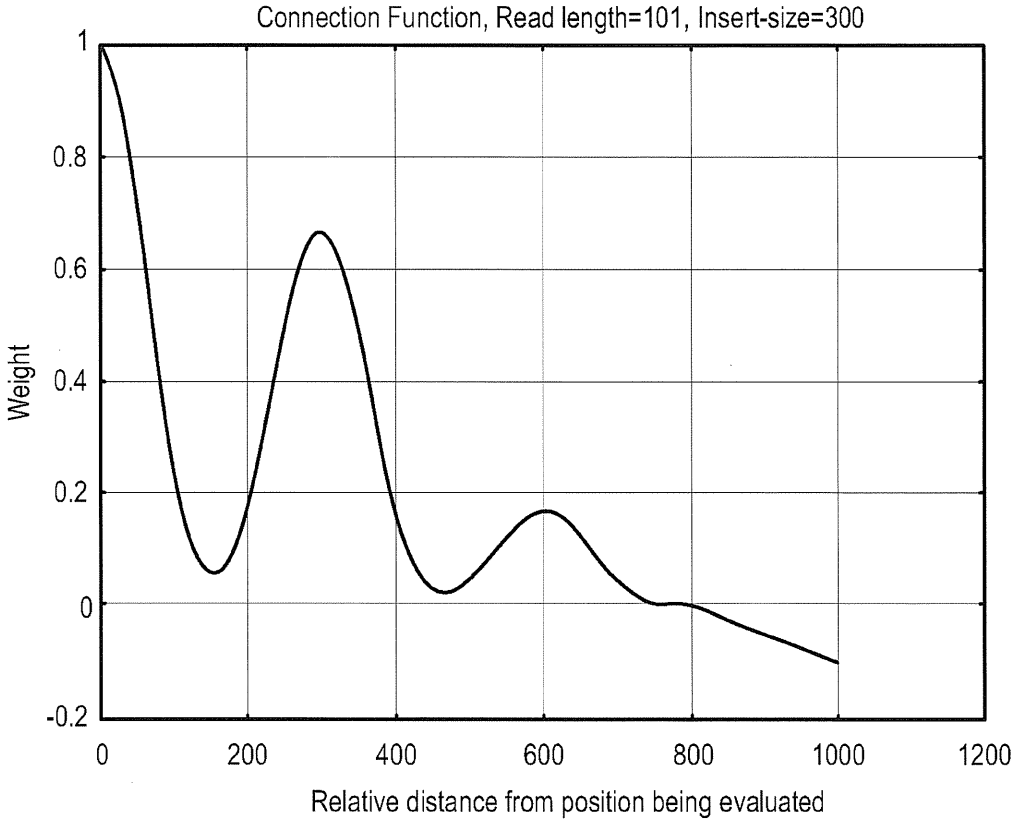


FIG. 23

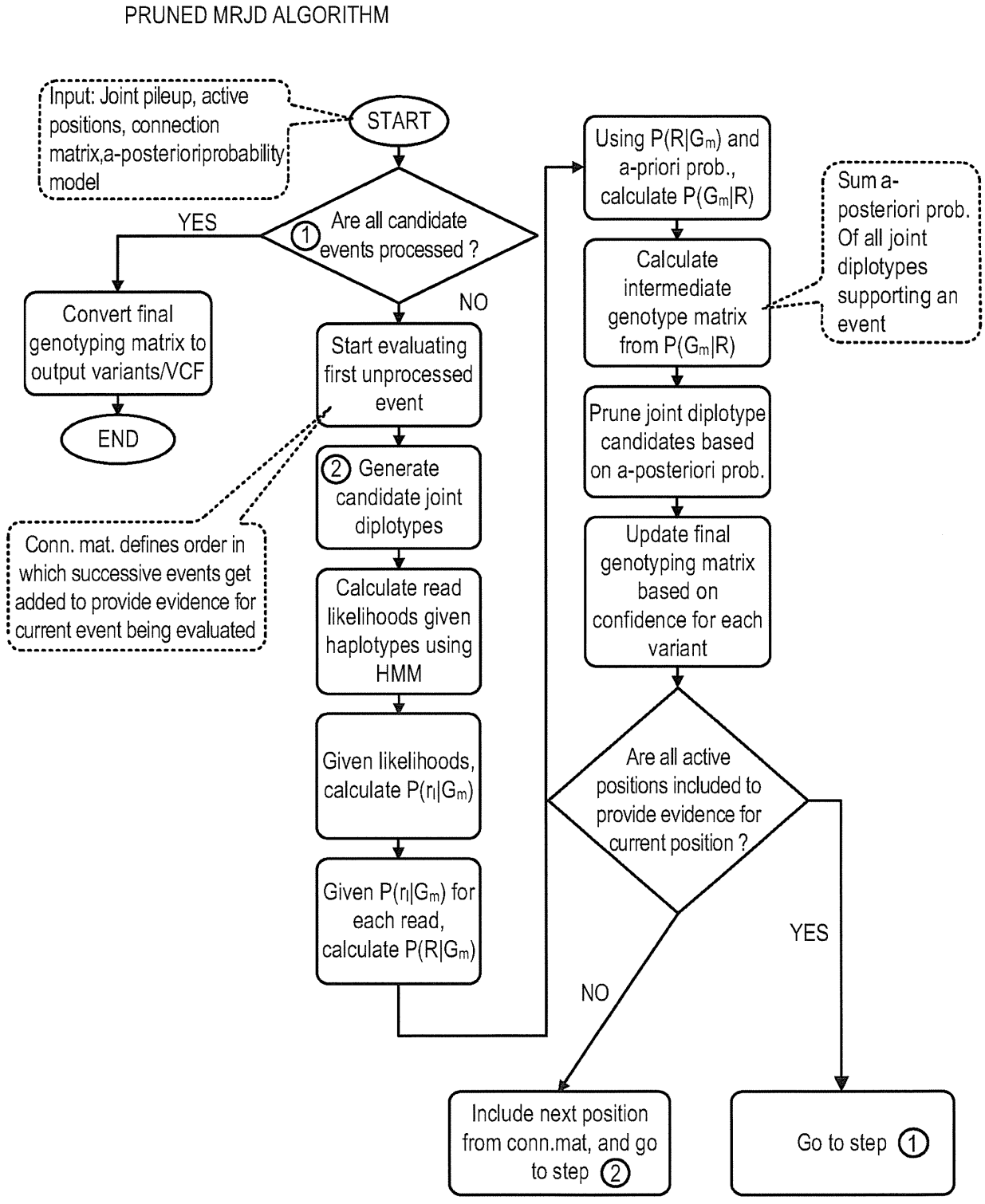

PRUNED MRJD ALGORITHM
Input: Joint pileup, active positions, connection matrix,a-posterioriprobability model
START
① Are all candidate events processed ?
YES
NO
Convert final genotyping matrix to output variants/VCF
END
Start evaluating first unprocessed event
② Generate candidate joint diplotypes
Calculate read likelihoods given haplotypes using HMM
Given likelihoods, calculate P(rl|Gm)
Given P(rl|Gm) for each read, calculate P(R|Gm)
Conn. mat. defines order in which successive events get added to provide evidence for current event being evaluated
Using P(R|Gm) and a-priori prob., calculate P(Gm|R)
Calculate intermediate genotype matrix from P(Gm|R)
Sum a-posteriori prob. Of all joint diplotypes supporting an event
Prune joint diplotype candidates based on a-posteriori prob.
Update final genotyping matrix based on confidence for each variant
Are all active positions included to provide evidence for current position ?
YES
NO
Include next position from conn.mat, and go to step ②
Go to step ①

FIG. 24

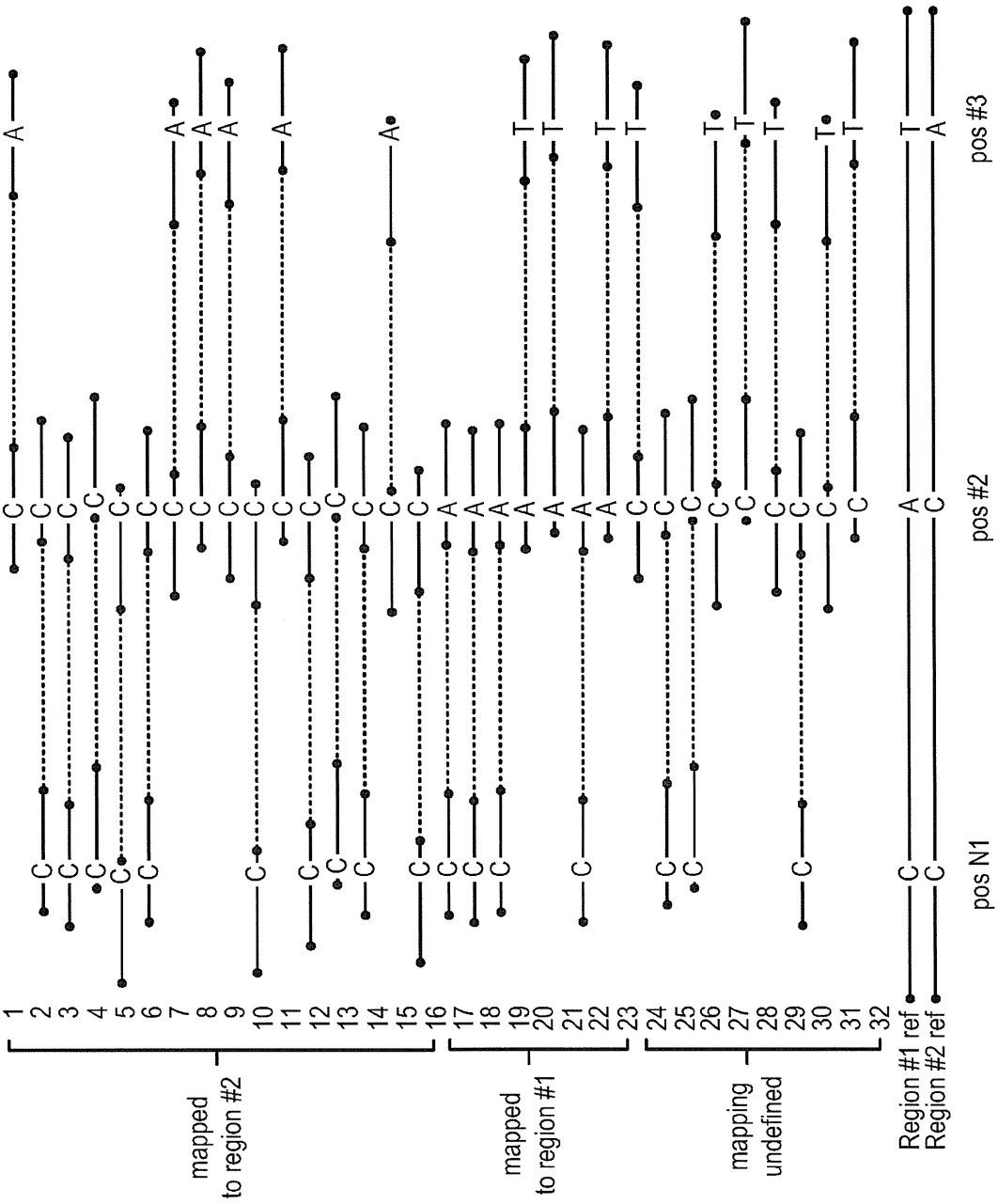

mapped to region #2
mapped to region #1
mapping undefined
Region #1 ref
Region #2 ref
pos N1
pos #2
pos #3

FIG. 25

i	H_1=CAT	H_2=CAA	H_3=CCT	H_4=CCA	H_5=GAT	H_6=GAA	H_7=GCT	H_8=GCA	CAT GCA CAT GCA	CAT GCA CCT GCA
1	0.0000	0.0033	0.0033	0.9801	0.0000	0.0033	0.0033	0.9801	0.4901	0.4909
2	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
3	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
4	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
5	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
6	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
7	0.0000	0.0033	0.0033	0.9801	0.0000	0.0033	0.0033	0.9801	0.4901	0.4909
8	0.0000	0.0033	0.0033	0.9801	0.0000	0.0033	0.0033	0.9801	0.4901	0.4909
9	0.0000	0.0033	0.0033	0.9801	0.0000	0.0033	0.0033	0.9801	0.4901	0.4909
10	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
11	0.0000	0.0033	0.0033	0.9801	0.0000	0.0033	0.0033	0.9801	0.4901	0.4909
12	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
13	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
14	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
15	0.0000	0.0033	0.0033	0.9801	0.0000	0.0033	0.0033	0.9801	0.4901	0.4909
16	0.0000	0.0000	0.0033	0.0033	0.0033	0.0033	0.9801	0.9801	0.4901	0.4909
17	0.9801	0.9801	0.0033	0.0033	0.0033	0.0033	0.0000	0.0000	0.4901	0.2459
18	0.9801	0.9801	0.0033	0.0033	0.0033	0.0033	0.0000	0.0000	0.4901	0.2459
19	0.9801	0.9801	0.0033	0.0033	0.0033	0.0033	0.0000	0.0000	0.4901	0.2459
20	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.4901	0.2459
21	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.4901	0.2459
22	0.9801	0.9801	0.0033	0.0033	0.0033	0.0033	0.0000	0.0000	0.4901	0.2459
23	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.4901	0.2459
24	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.2475
25	0.0033	0.0033	0.9801	0.9801	0.0000	0.0000	0.0033	0.0033	0.0033	0.2475
26	0.0033	0.0033	0.9801	0.9801	0.0000	0.0000	0.0033	0.0033	0.0033	0.2475
27	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.2475
28	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.2475
29	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.2475
30	0.0033	0.0033	0.9801	0.9801	0.0000	0.0000	0.0033	0.0033	0.0033	0.2475
31	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.2475
32	0.0033	0.0000	0.9801	0.0033	0.0033	0.0000	0.9801	0.0033	0.0033	0.2475
									3.5e-30	2.2e-15

FIG. 26

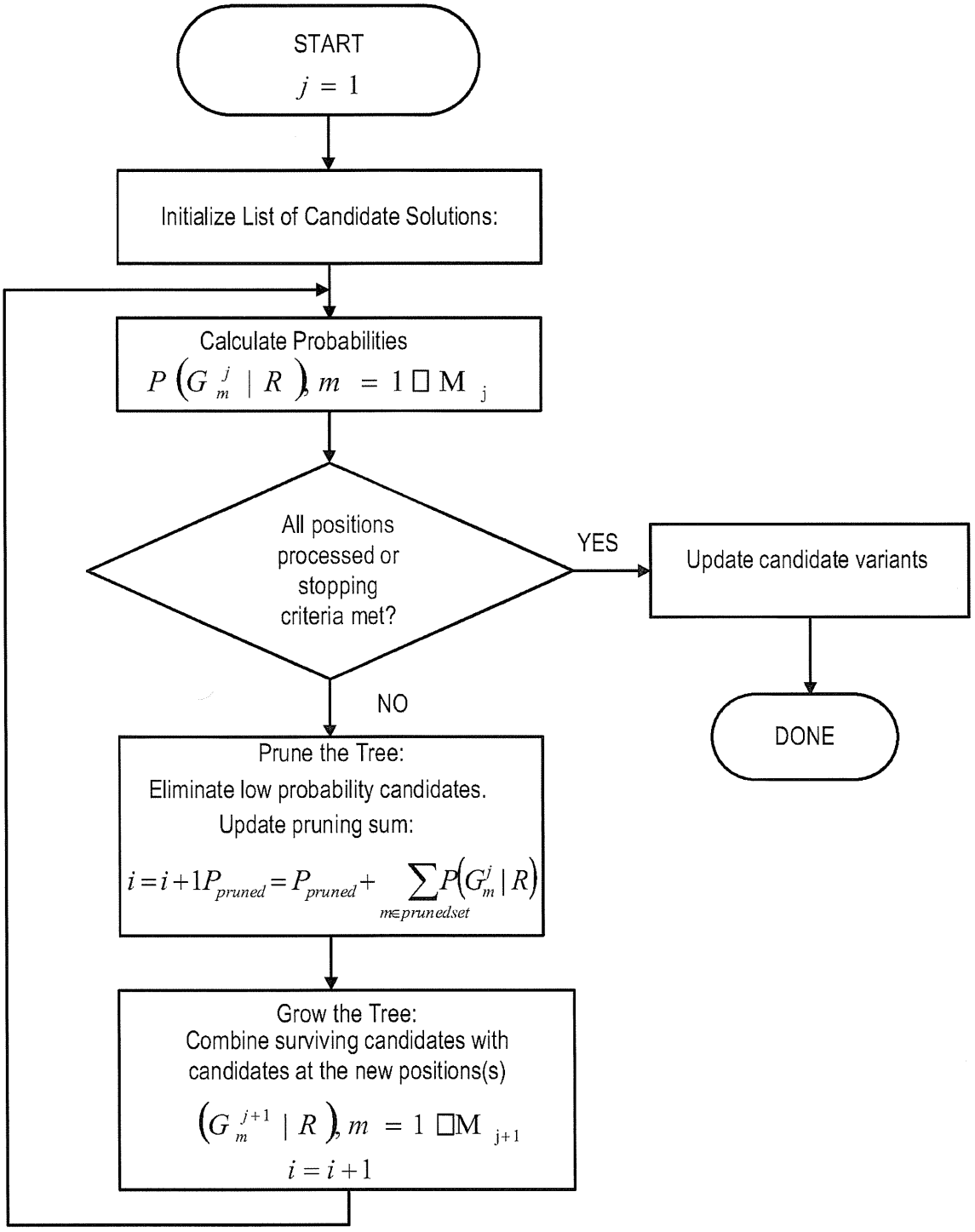

START
$j = 1$
Initialize List of Candidate Solutions:
Calculate Probabilities
$P\left(G_m^j \mid R\right), m = 1 \square M_j$
All positions processed or stopping criteria met?
YES
Update candidate variants
DONE
NO
Prune the Tree:
Eliminate low probability candidates.
Update pruning sum:
$i = i+1 P_{pruned} = P_{pruned} + \sum_{m \in prunedset} P\left(G_m^j \mid R\right)$
Grow the Tree:
Combine surviving candidates with candidates at the new positions(s)
$\left(G_m^{j+1} \mid R\right), m = 1 \square M_{j+1}$
$i = i+1$

FIG. 27

$G^1_1 = \begin{bmatrix} C & G \\ C & G \end{bmatrix}$	$P(G_1\|R) = 0.99892171$ (phred 0.00)
$G^1_2 = \begin{bmatrix} G & G \\ C & G \end{bmatrix}$	$P(G_2\|R) = 0.00103398$ (phred 29.85)
$G^1_3 = \begin{bmatrix} C & G \\ C & C \end{bmatrix}$	$P(G_3\|R) = 0.00003934$ (phred 44.05)
$G^1_4 = \begin{bmatrix} G & G \\ C & C \end{bmatrix}$	$P(G_4\|R) = 0.00000251$ (phred 56.00)
$G^1_5 = \begin{bmatrix} G & C \\ G & C \end{bmatrix}$	$P(G_5\|R) = 0.00000126$ (phred 59.00)
$G^1_6 = \begin{bmatrix} G & G \\ G & C \end{bmatrix}$	$P(G_6\|R) = 0.00000116$ (phred 59.35)
$G^1_7 = \begin{bmatrix} G & C \\ C & C \end{bmatrix}$	$P(G_7\|R) = 0.00000004$ (phred 73.55)
$G^1_8 = \begin{bmatrix} G & G \\ G & G \end{bmatrix}$	$P(G_8\|R) = 0.00000000$ (phred 151.67)
$G^1_8 = \begin{bmatrix} C & C \\ C & C \end{bmatrix}$	$P(G_9\|R) = 0.00000000$ (phred 225.85)

FIG. 28

G_1^2 =	$\begin{bmatrix} CC & GC \\ CA & GC \end{bmatrix}$	$P(G_1\|R) = 0.99799456$ (phred 0.01)
G_2^2 =	$\begin{bmatrix} CA & GC \\ CA & GC \end{bmatrix}$	$P(G_2\|R) = 0.00101673$ (phred 29.93)
G_3^2 =	$\begin{bmatrix} CA & GC \\ CA & CC \end{bmatrix}$	$P(G_3\|R) = 0.00097796$ (phred 30.10)
G_4^2 =	$\begin{bmatrix} CC & GC \\ CA & CC \end{bmatrix}$	$P(G_4\|R) = 0.00000913$ (phred 50.39)
G_5^2 =	$\begin{bmatrix} GC & GC \\ CA & CC \end{bmatrix}$	$P(G_5\|R) = 0.00000125$ (phred 59.02)
G_6^2 =	$\begin{bmatrix} GC & GC \\ CA & GC \end{bmatrix}$	$P(G_6\|R) = 0.00000015$ (phred 68.29)
G_7^2 =	$\begin{bmatrix} GA & GC \\ CA & CC \end{bmatrix}$	$P(G_7\|R) = 0.00000007$ (phred 71.45)
G_8^2 =	$\begin{bmatrix} CC & GC \\ CA & GA \end{bmatrix}$	$P(G_8\|R) = 0.00000007$ (phred 71.45)
G_9^2 =	$\begin{bmatrix} GA & GC \\ CA & GC \end{bmatrix}$	$P(G_9\|R) = 0.00000006$ (phred 72.06)
G_{10}^2 =	$\begin{bmatrix} CC & GC \\ CC & CA \end{bmatrix}$	$P(G_{10}\|R) = 0.00000001$ (phred 79.89)
G_{11}^2 =	$\begin{bmatrix} GC & CC \\ GC & CA \end{bmatrix}$	$P(G_{11}\|R) = 0.00000000$ (phred 88.51)
G_{12}^2 =	$\begin{bmatrix} GC & GC \\ CC & CA \end{bmatrix}$	$P(G_{12}\|R) = 0.00000000$ (phred 88.51)
G_{13}^2 =	$\begin{bmatrix} CC & GC \\ CA & CA \end{bmatrix}$	$P(G_{13}\|R) = 0.00000000$ (phred 89.10)
G_{14}^2 =	$\begin{bmatrix} GA & GC \\ CC & GC \end{bmatrix}$	$P(G_{14}\|R) = 0.00000000$ (phred 116.35)
G_{15}^2 =	$\begin{bmatrix} CA & GC \\ CA & GA \end{bmatrix}$	$P(G_{15}\|R) = 0.00000000$ (phred 120.60)

(candidates 16-70 not shown)

FIG. 29

G^3_1 =	CCT GCA CAT GCA	$P(G_1\|R) = 0.99885810$ (phred 0.00)
G^3_2 =	CCT GCT CAT GCA	$P(G_2\|R) = 0.00111991$ (phred 29.51)
G^3_3 =	CCT GCA CAT CCA	$P(G_3\|R) = 0.00000914$ (phred 50.39)
G^3_4 =	CAT GCT CAT GCA	$P(G_4\|R) = 0.00000462$ (phred 53.36)
G^3_5 =	CAT GCA CAT CCT	$P(G_5\|R) = 0.00000222$ (phred 56.53)
G^3_6 =	CAT GCT CAT CCA	$P(G_6\|R) = 0.00000222$ (phred 56.53)
G^3_7 =	CCA GCT CAT GCT	$P(G_7\|R) = 0.00000126$ (phred 59.01)
G^3_8 =	GCT GCA CAT CCA	$P(G_8\|R) = 0.00000126$ (phred 59.01)
G^3_9 =	CCA GCT CAT GCA	$P(G_9\|R) = 0.00000126$ (phred 59.01)
G^3_{10} =	CCT GCA CAT CCT	$P(G_{10}\|R) = 0.00000001$ (phred 79.89)
G^3_{11} =	CCT GCT CAT CCA	$P(G_{11}\|R) = 0.00000001$ (phred 79.89)
G^3_{12} =	GCT GCA CAT CCT	$P(G_{12}\|R) = 0.00000000$ (phred 88.52)
G^3_{13} =	GCT GCT CAT CCA	$P(G_{13}\|R) = 0.00000000$ (phred 88.52)
G^3_{14} =	CAT GCT CAA GCA	$P(G_{14}\|R) = 0.00000000$ (phred 91.82)
G^3_{15} =	CCT GCA CAA GCA	$P(G_{15}\|R) = 0.00000000$ (phred 94.70)

(candidates 16-65 not shown)

FIG. 30

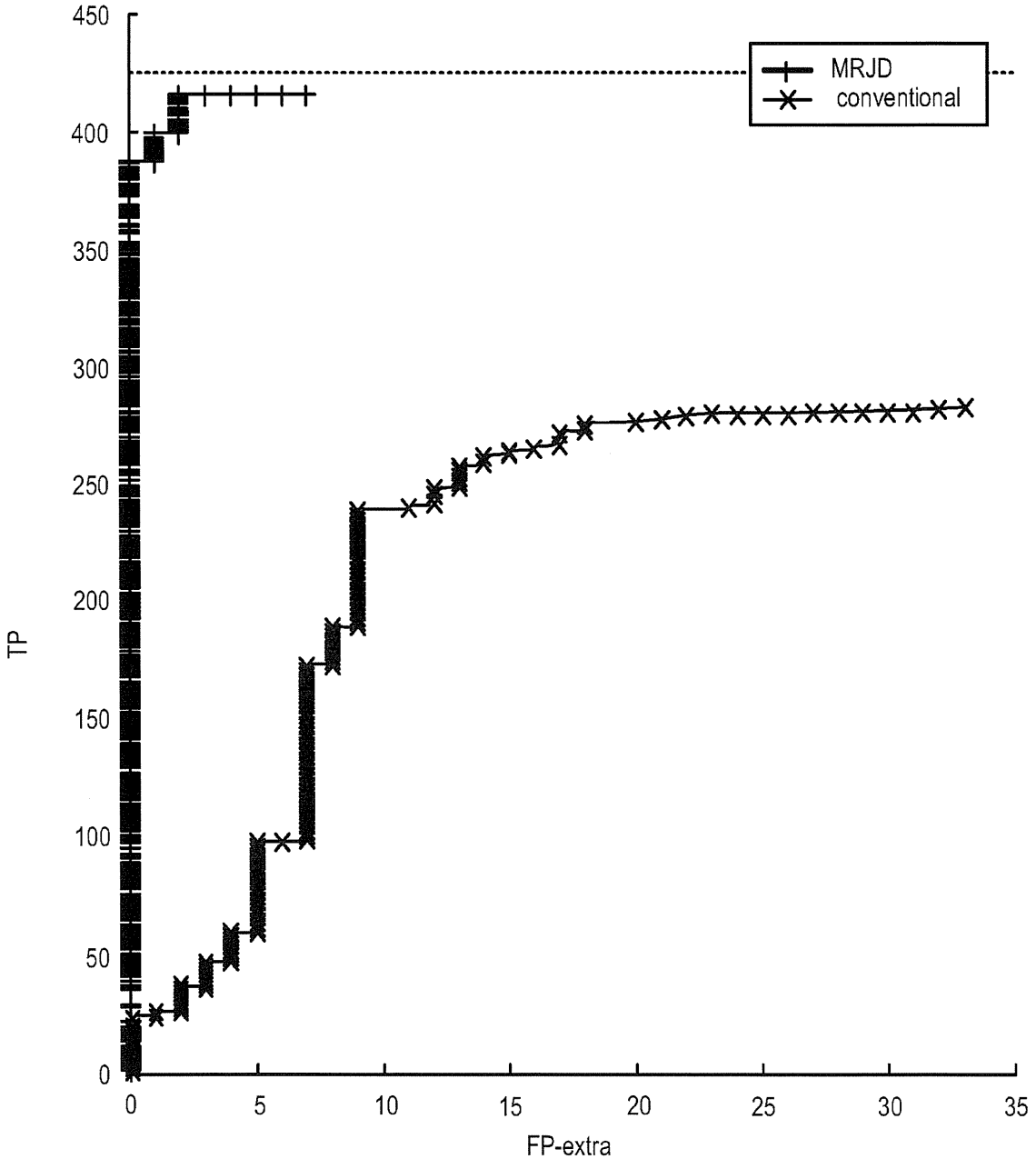

450
400
350
300
250
200
150
100
50
0
TP
0
5
10
15
20
25
30
35
FP-extra
MRJD
conventional

FIG. 31

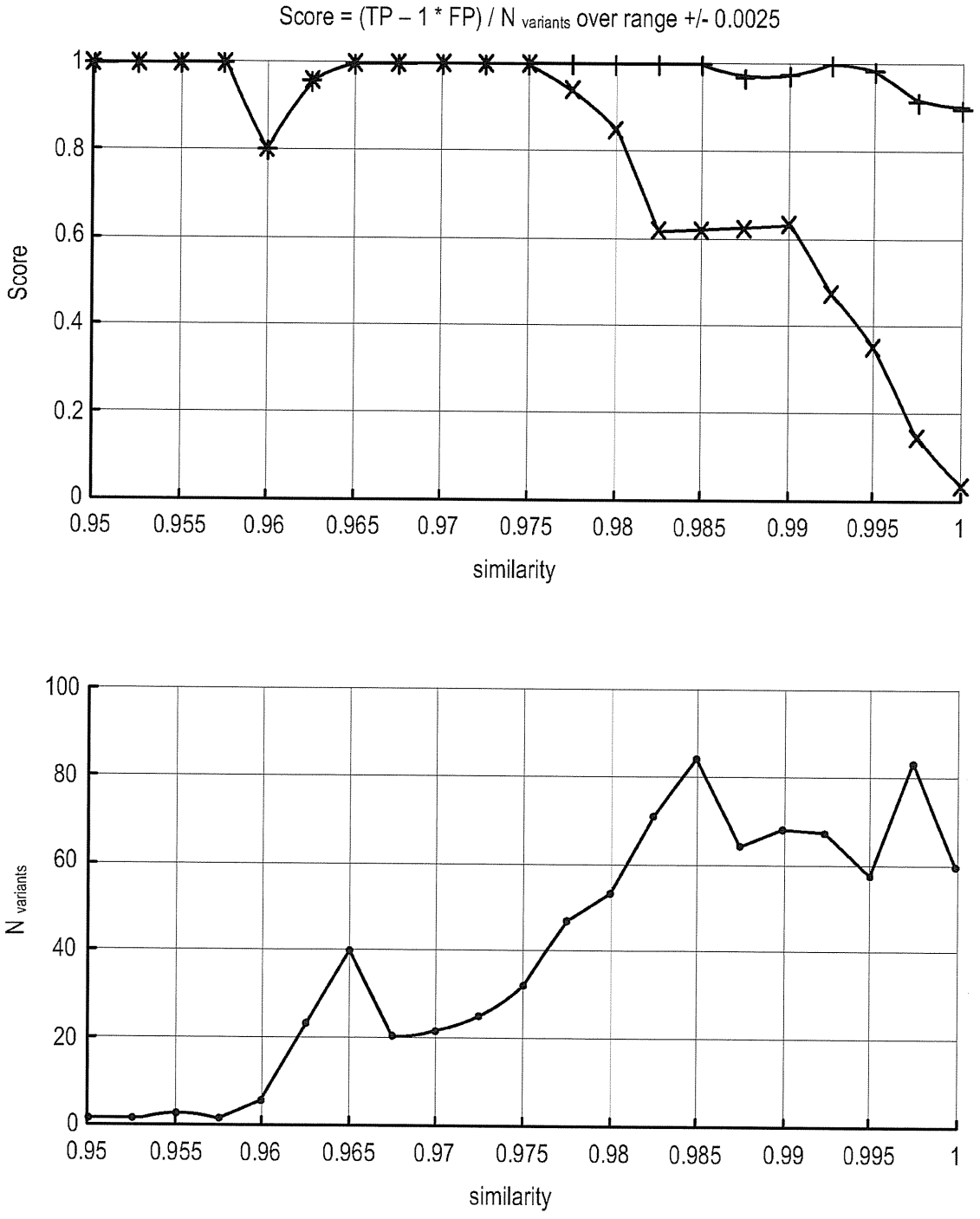

Score = (TP – 1 * FP) / N variants over range +/- 0.0025
Score
1
0.8
0.6
0.4
0.2
0
0.95
0.955
0.96
0.965
0.97
0.975
0.98
0.985
0.99
0.995
1
similarity
N variants
100
80
60
40
20
0
0.95
0.955
0.96
0.965
0.97
0.975
0.98
0.985
0.99
0.995
1
similarity

FIG. 32

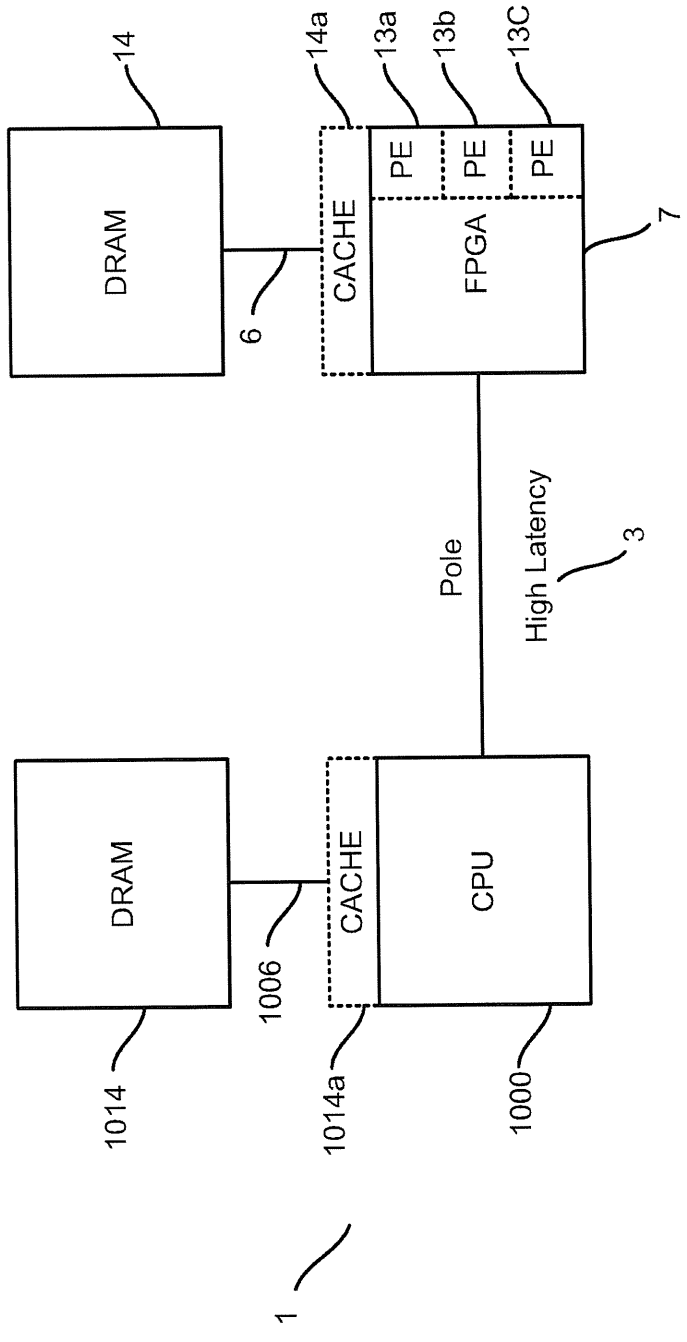

1
1014
1014a
1006
1000
DRAM
CACHE
CPU
Pole
High Latency
3
DRAM
14
6
14a
CACHE
13a
13b
13C
PE
PE
PE
FPGA
7

FIG. 33A

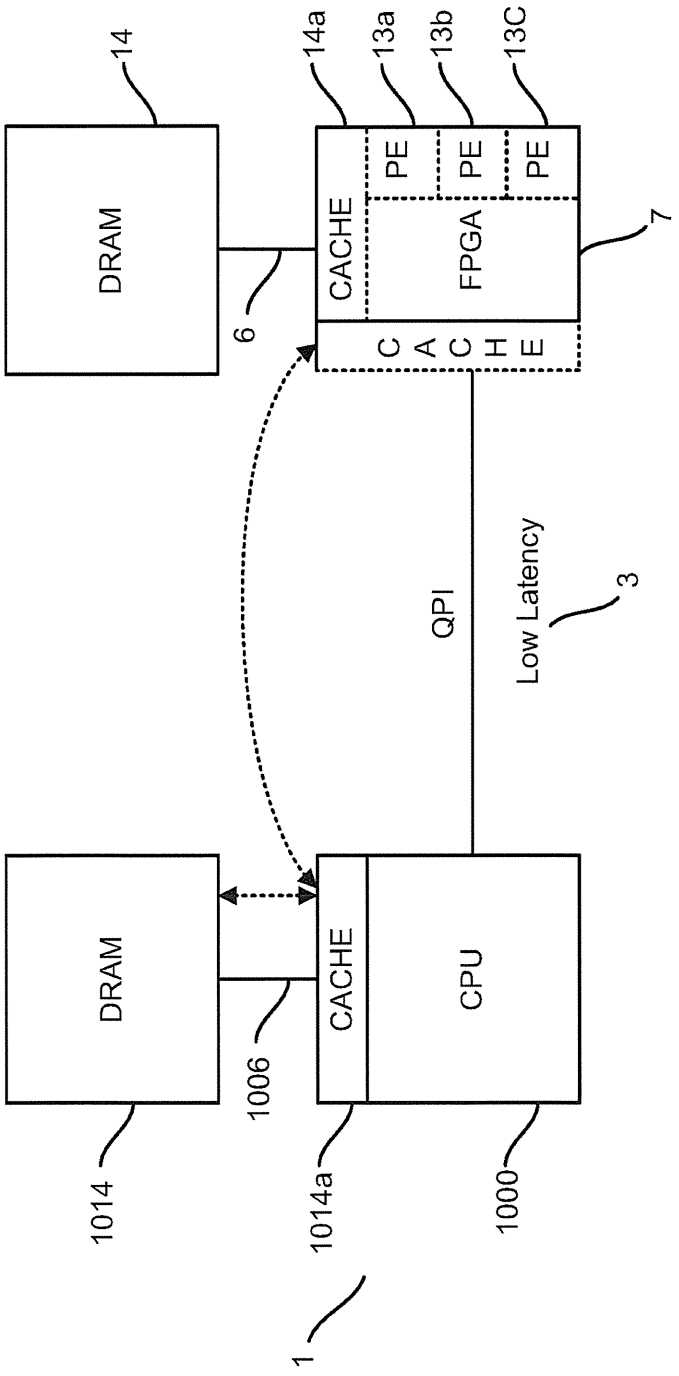

14
14a
13a
13b
13C
DRAM
CACHE
PE
PE
PE
FPGA
7
6
C A C H E
QPI
Low Latency
3
DRAM
CACHE
CPU
1006
1014
1014a
1000
1

FIG. 33B

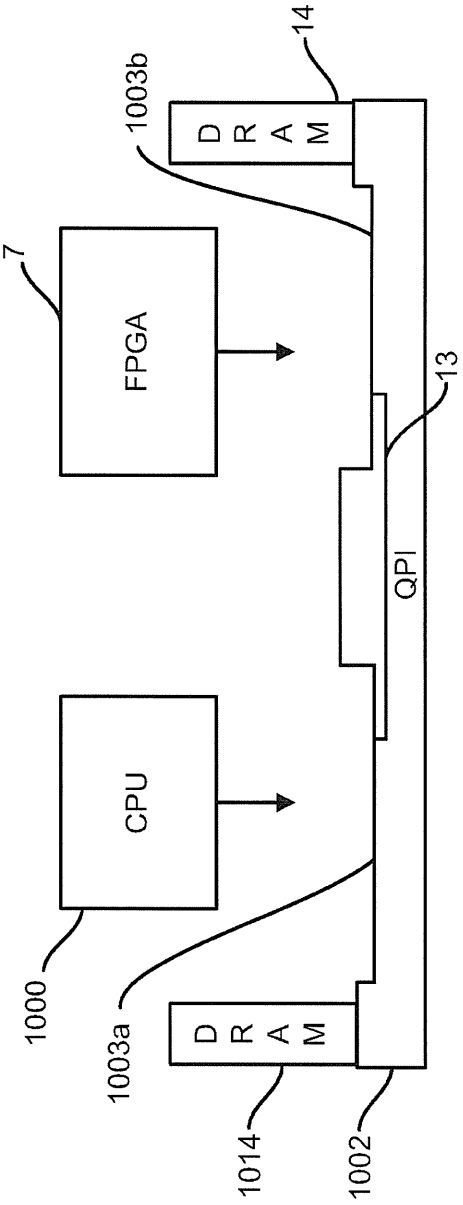


FIG. 34A

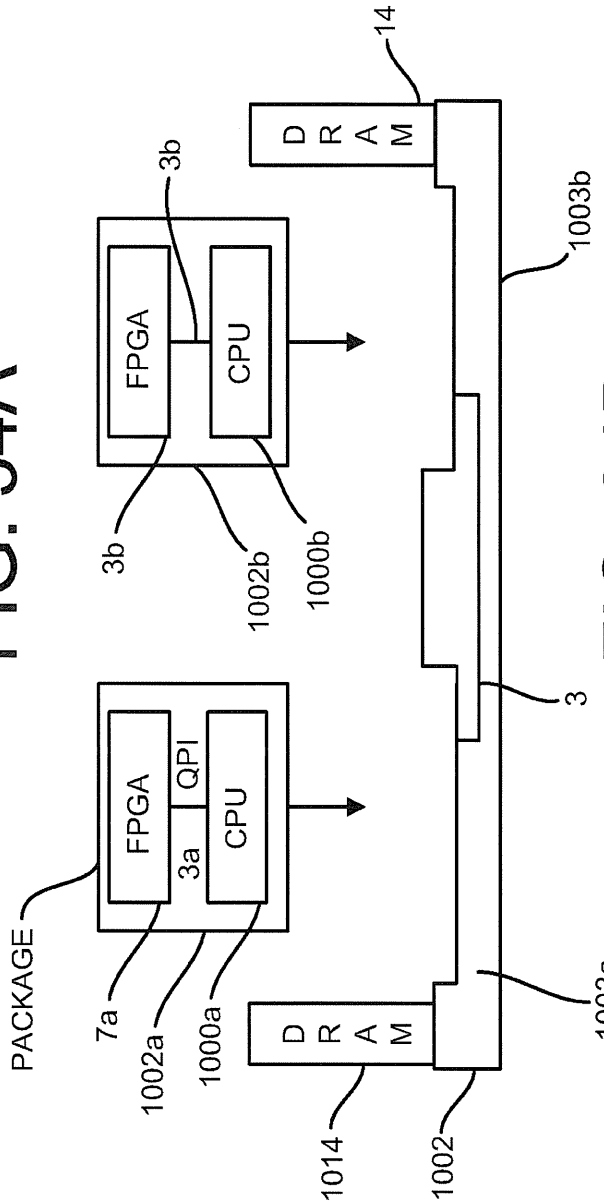


FIG. 34B

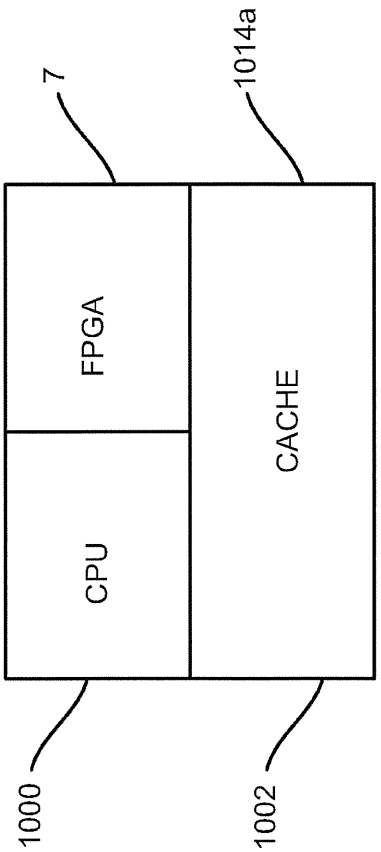


FIG. 35

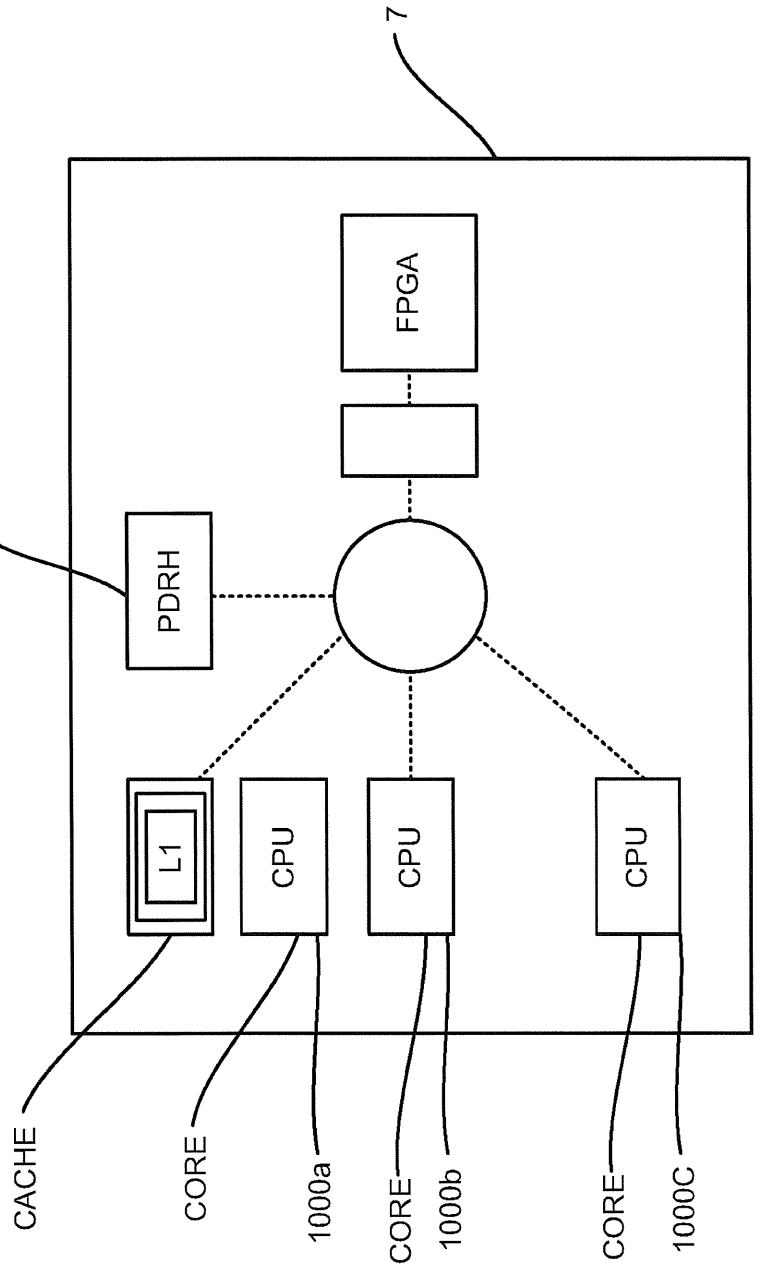

7
FPGA
1014
PDRH
L1
CPU
CPU
CPU
CACHE
CORE
1000a
CORE
1000b
CORE
1000C

FIG. 36

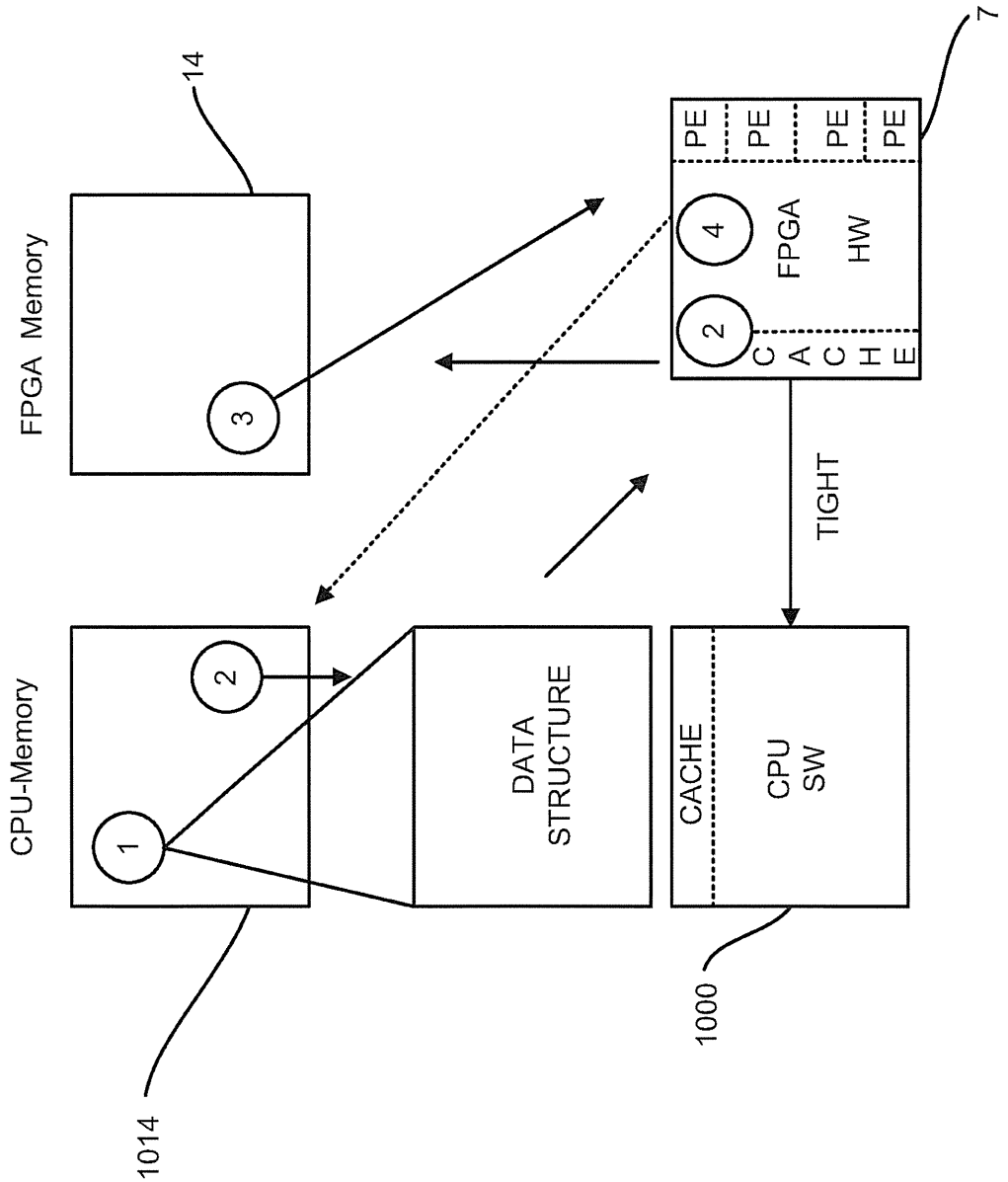

14
FPGA Memory
3
PE
PE
PE
PE
4
FPGA
HW
2
CACHE
7
TIGHT
CPU-Memory
1
2
DATA
STRUCTURE
CACHE
CPU
SW
1014
1000

FIG. 37

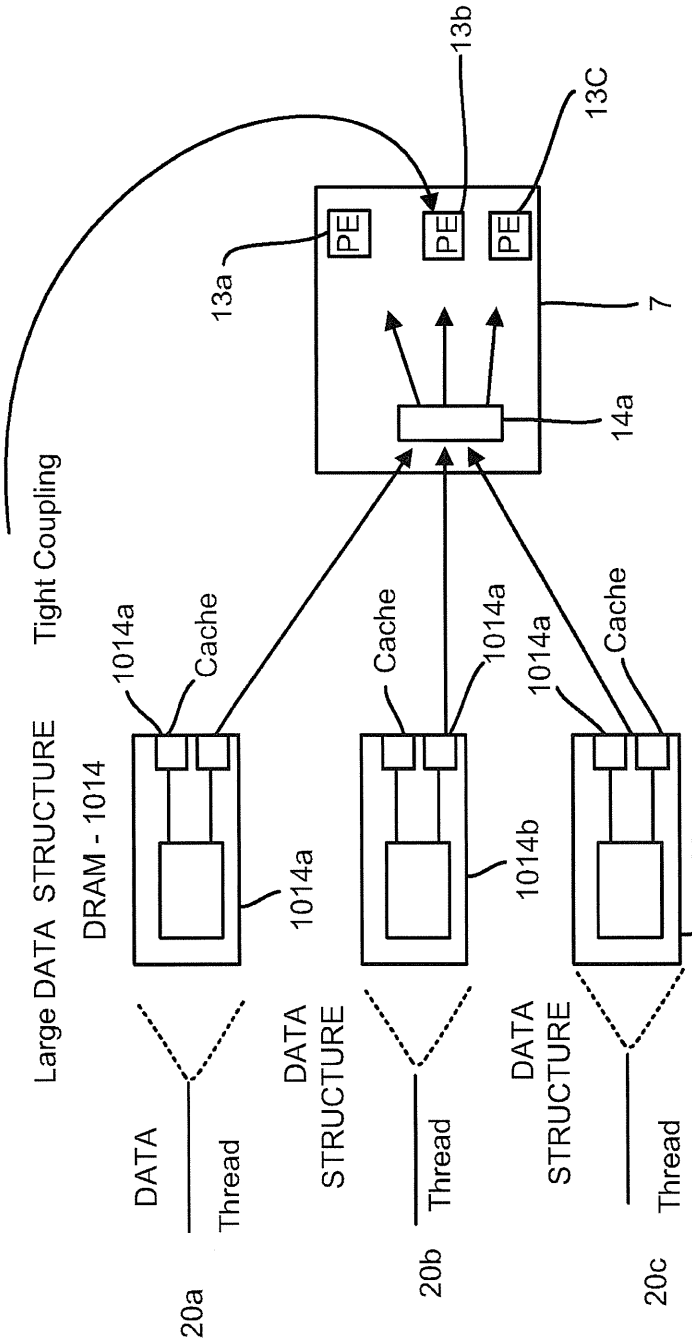

Large DATA STRUCTURE
DRAM - 1014
Tight Coupling
1014a
Cache
1014a
1014b
1014C
Memory
Working space
3
DATA
STRUCTURE
Thread
20a
20b
20c
PE
PE
PE
13a
13b
13C
7
14a
CPU
1000

FIG. 38

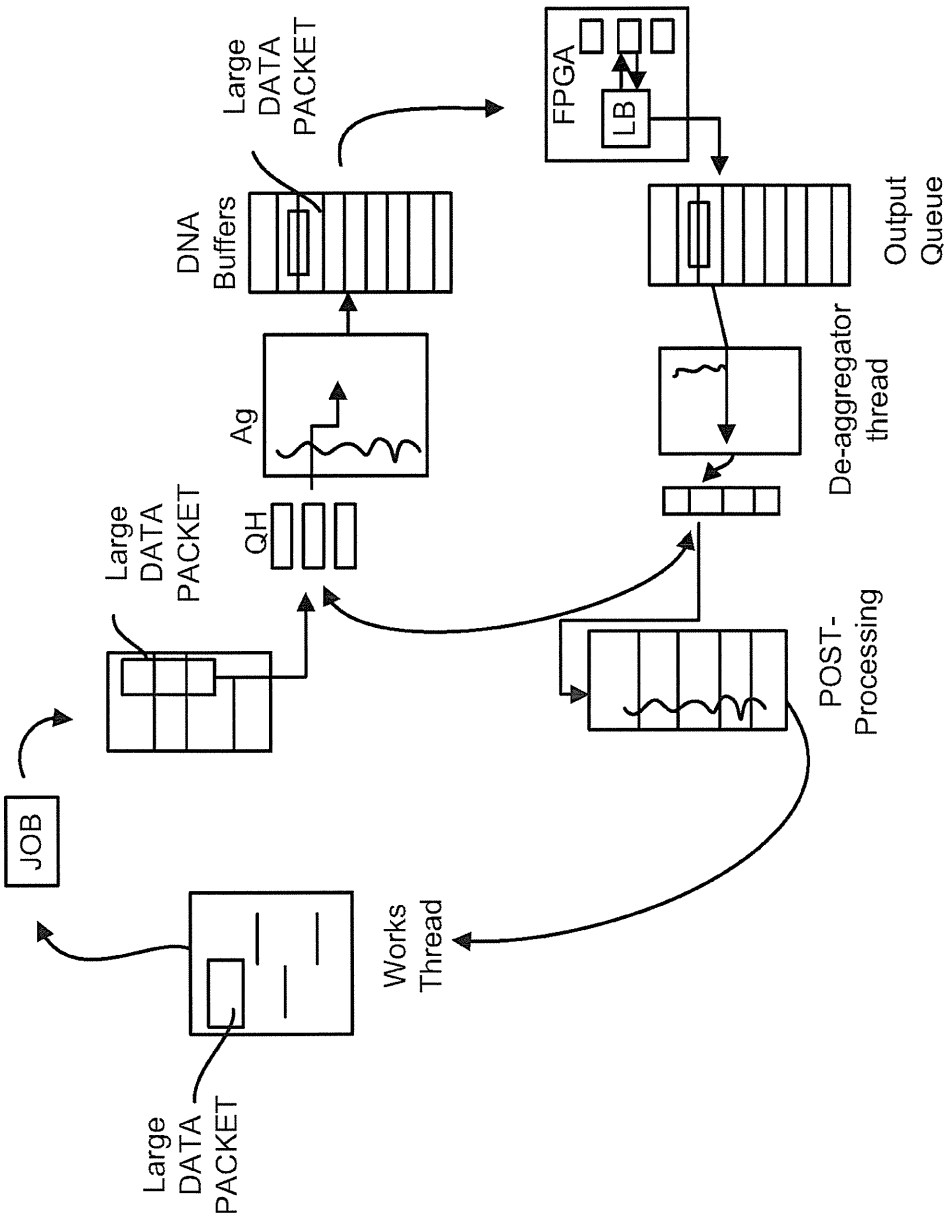

Large DATA PACKET
FPGA
LB
DNA Buffers
Output Queue
Ag
De-aggregator thread
Large DATA PACKET
QH
POST-Processing
JOB
Works Thread
Large DATA PACKET

FIG. 39

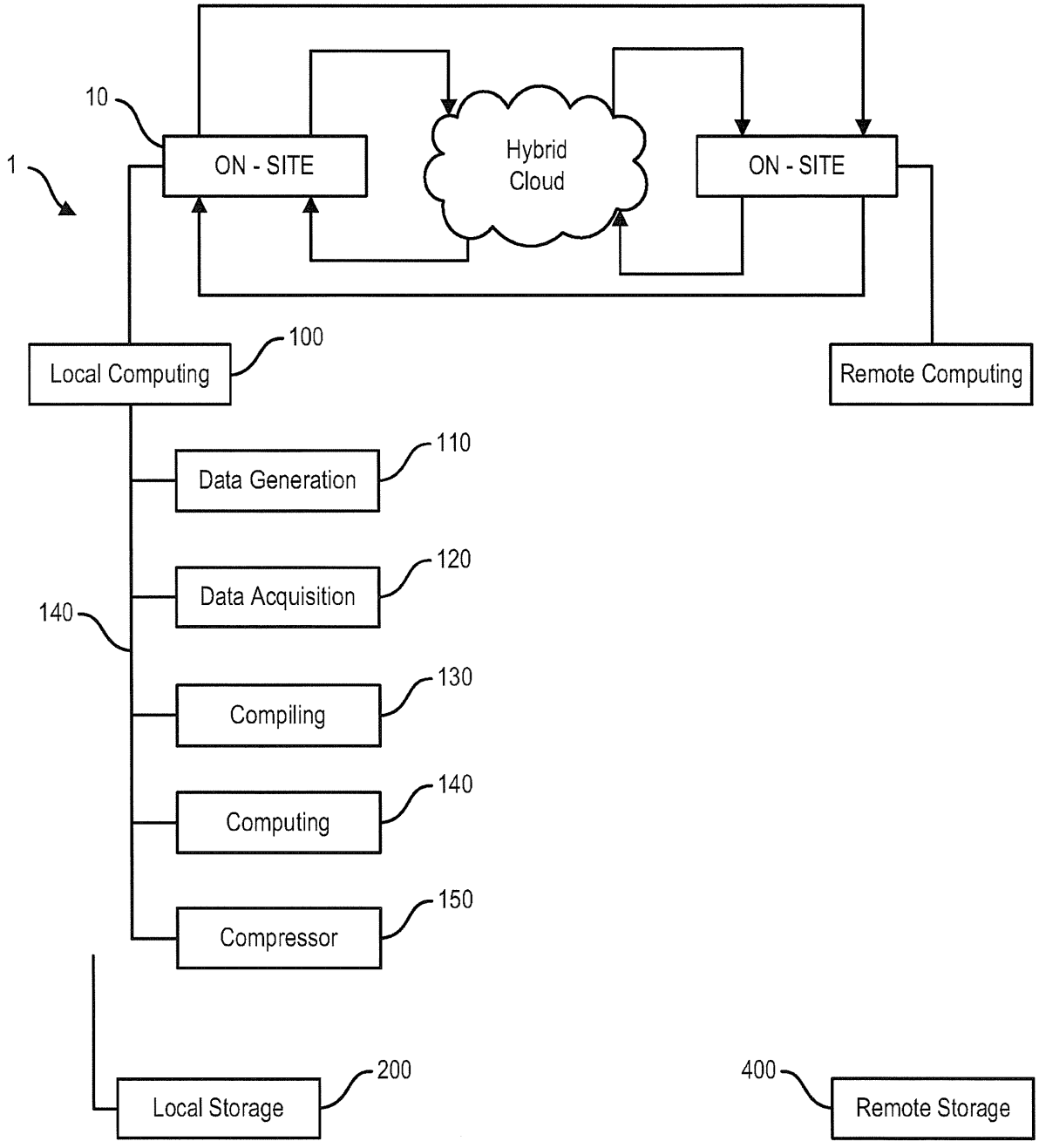

10
1
ON - SITE
Hybrid
Cloud
ON - SITE
100
Local Computing
Remote Computing
110
Data Generation
120
Data Acquisition
140
130
Compiling
140
Computing
150
Compressor
200
Local Storage
400
Remote Storage

FIG. 40A

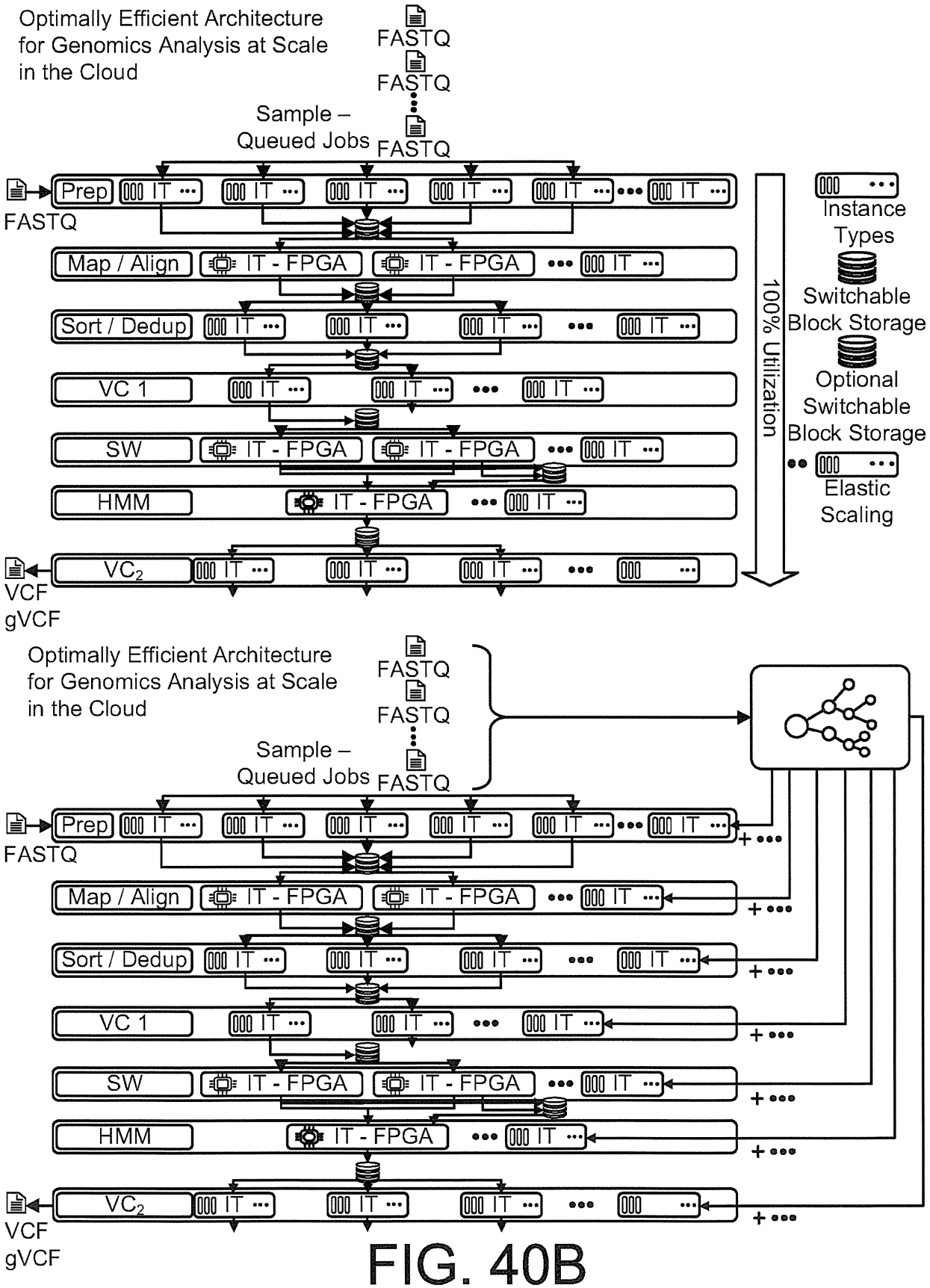

Optimally Efficient Architecture for Genomics Analysis at Scale in the Cloud
FASTQ
FASTQ
FASTQ
Sample – Queued Jobs
Prep
IT
FASTQ
Map / Align
IT - FPGA
Sort / Dedup
VC 1
SW
HMM
VC2
VCF
gVCF
100% Utilization
Instance Types
Switchable Block Storage
Optional Switchable Block Storage
Elastic Scaling
Optimally Efficient Architecture for Genomics Analysis at Scale in the Cloud
FASTQ
FASTQ
FASTQ
Sample – Queued Jobs
Prep
IT
FASTQ
Map / Align
IT - FPGA
Sort / Dedup
VC 1
SW
HMM
VC2
VCF
gVCF

FIG. 40B

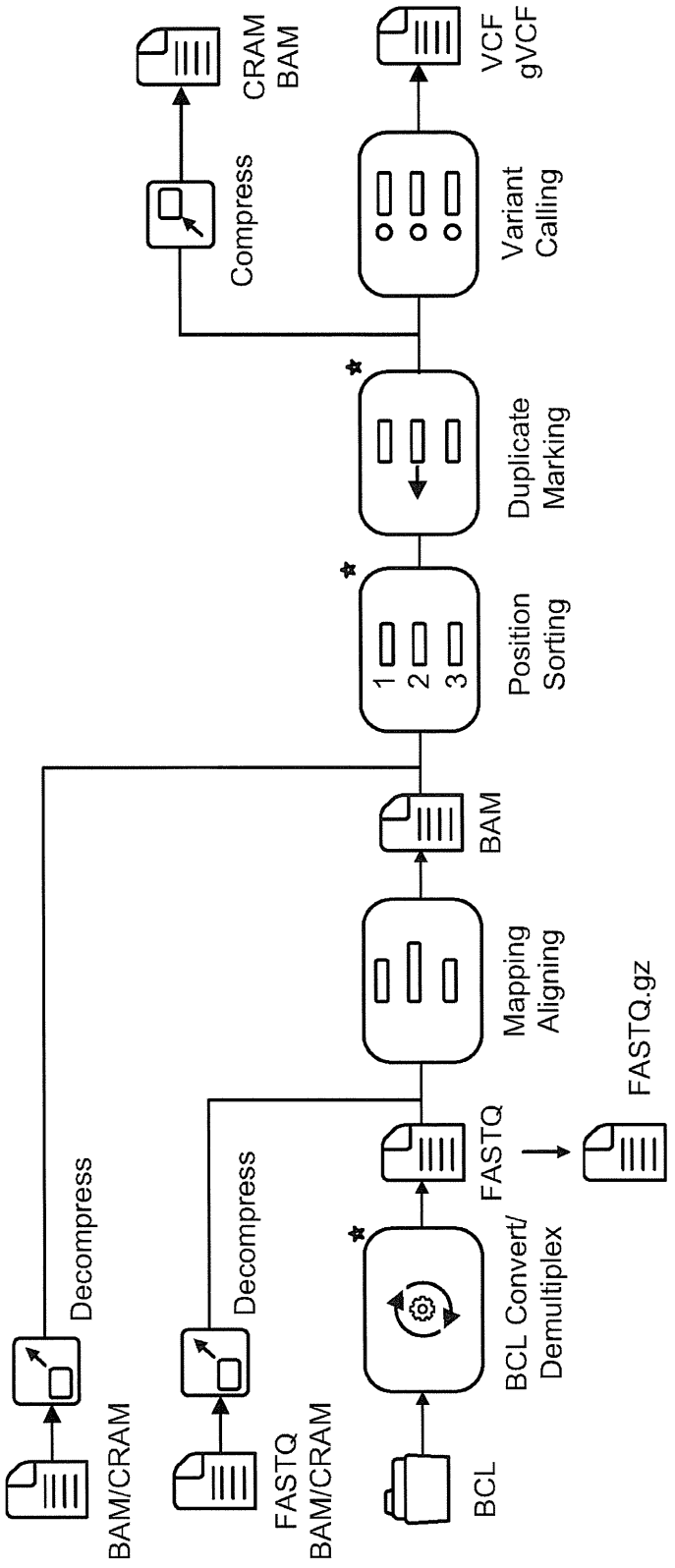

BAM/CRAM
Decompress
FASTQ
BAM/CRAM
Decompress
BCL
BCL Convert/
Demultiplex
FASTQ
FASTQ.gz
Mapping
Aligning
BAM
1
2
3
Position
Sorting
Duplicate
Marking
Variant
Calling
VCF
gVCF
Compress
CRAM
BAM

FIG. 40C

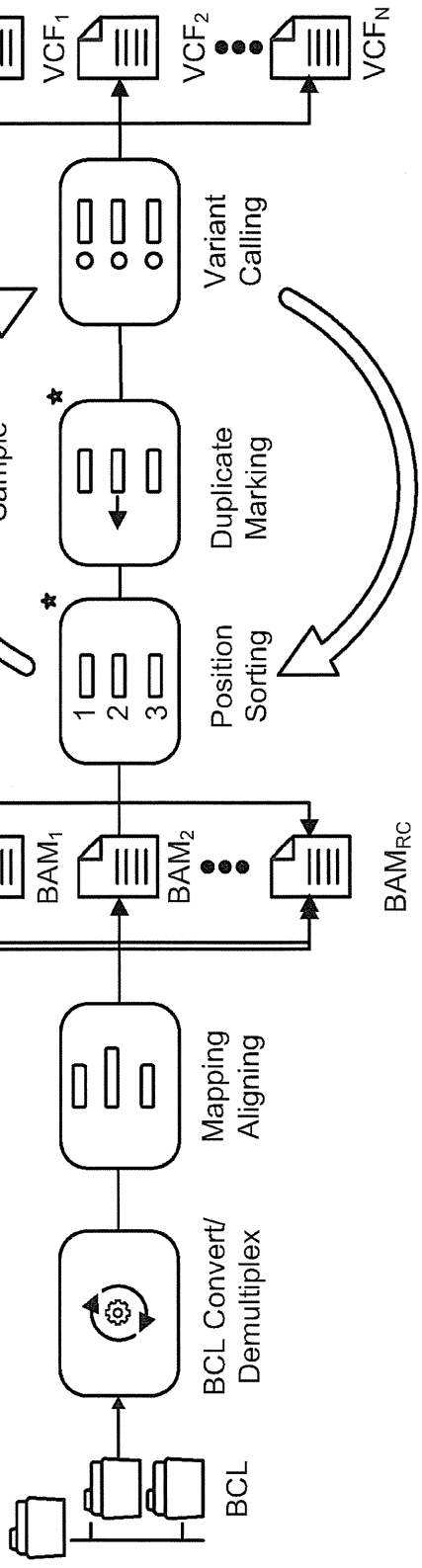

BCL
BCL Convert/ Demultiplex
Mapping Aligning
BAM1
BAM2
BAMRC
1
2
3
Position Sorting
Duplicate Marking
Variant Calling
Process by Sample
VCF1
VCF2
VCFN

FIG. 40D

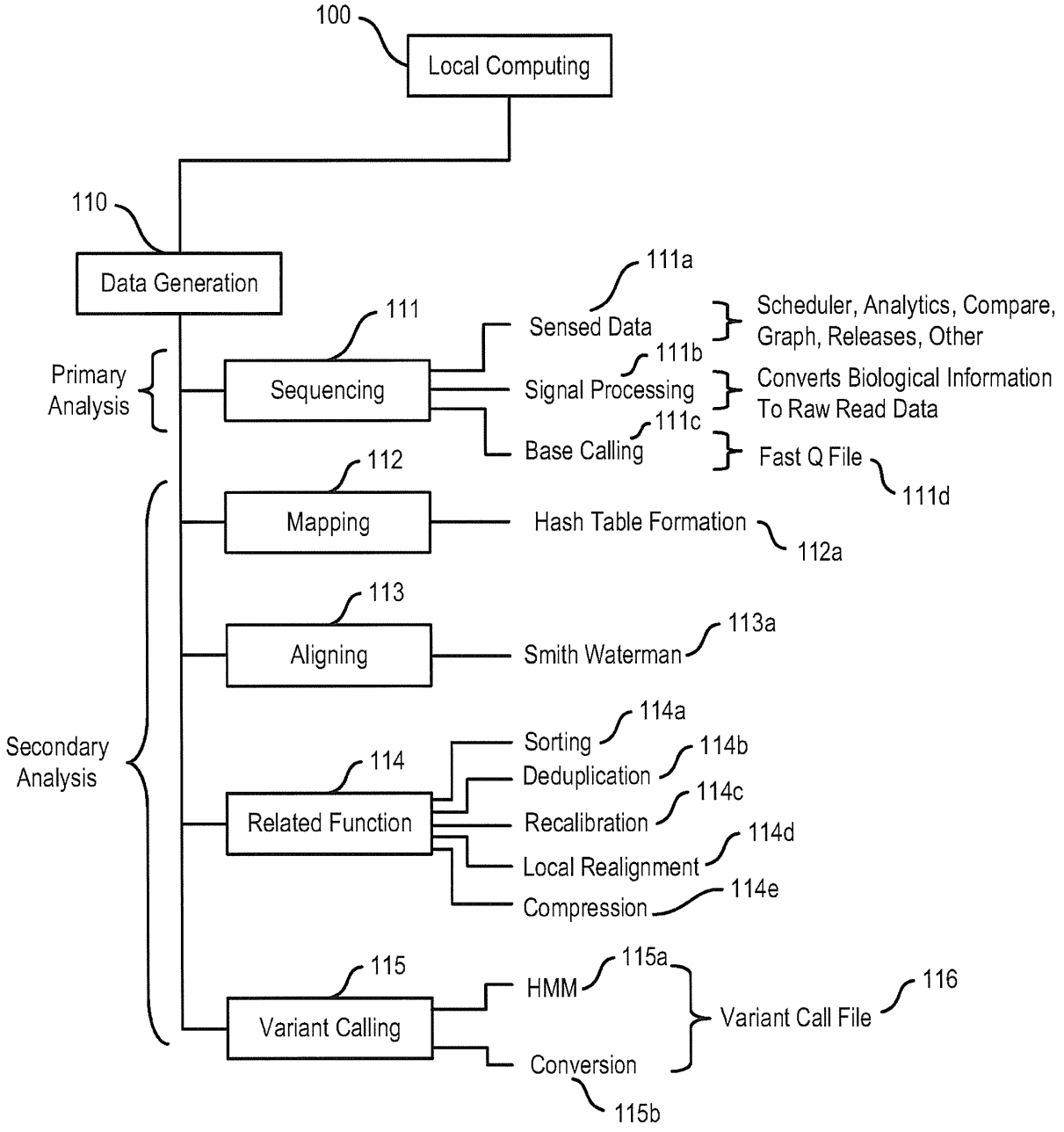

100
Local Computing
110
Data Generation
Primary Analysis
Secondary Analysis
111
Sequencing
111a
Sensed Data
Scheduler, Analytics, Compare, Graph, Releases, Other
111b
Signal Processing
Converts Biological Information To Raw Read Data
111c
Base Calling
Fast Q File
111d
112
Mapping
Hash Table Formation
112a
113
Aligning
Smith Waterman
113a
114
Related Function
114a
Sorting
114b
Deduplication
114c
Recalibration
114d
Local Realignment
114e
Compression
115
Variant Calling
115a
HMM
Conversion
115b
Variant Call File
116

FIG. 41A

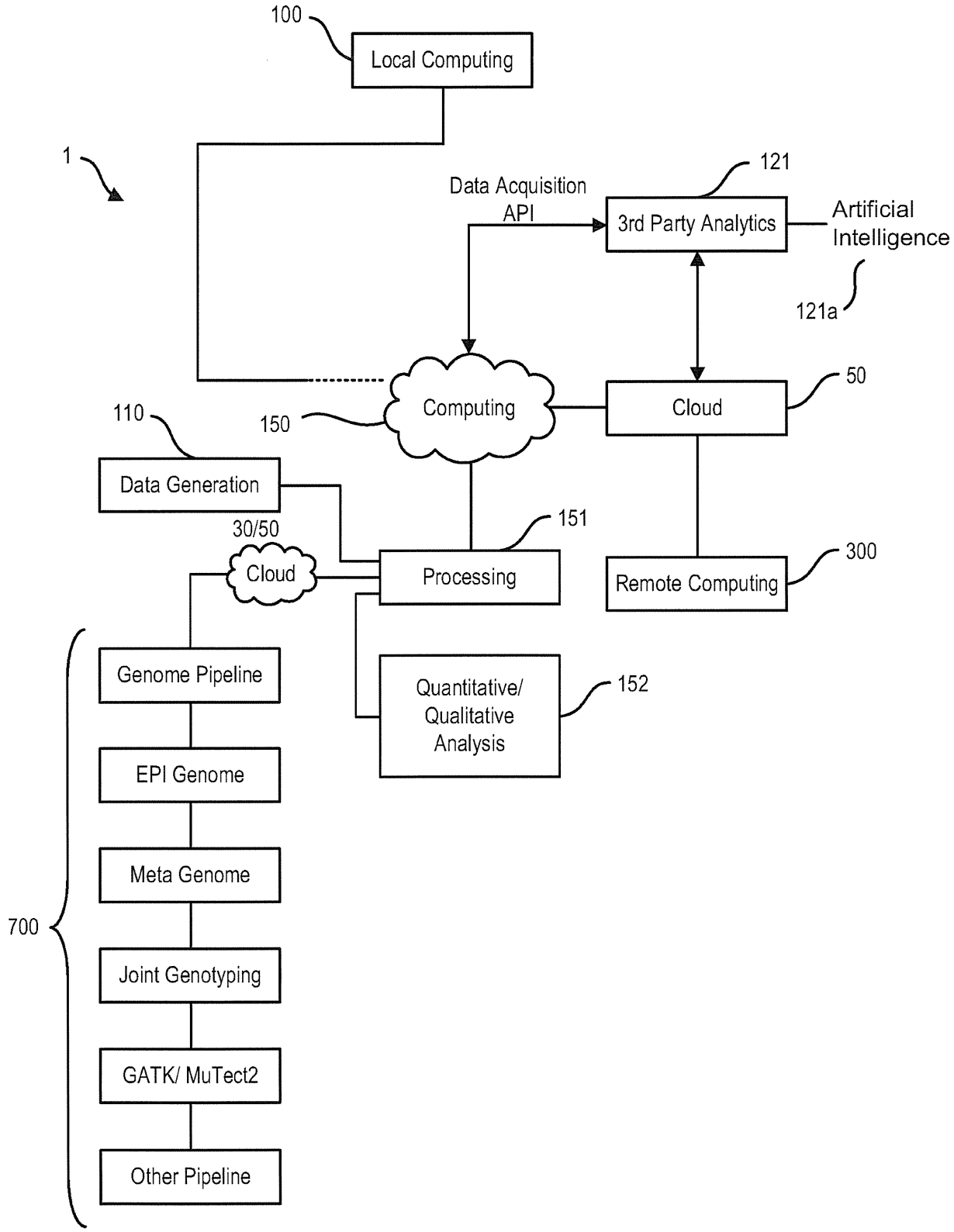

100
Local Computing
1
Data Acquisition API
121
3rd Party Analytics
Artificial Intelligence
121a
Computing
150
Cloud
50
110
Data Generation
30/50
Cloud
151
Processing
Remote Computing
300
Genome Pipeline
Quantitative/ Qualitative Analysis
152
EPI Genome
Meta Genome
700
Joint Genotyping
GATK/ MuTect2
Other Pipeline

FIG. 41B

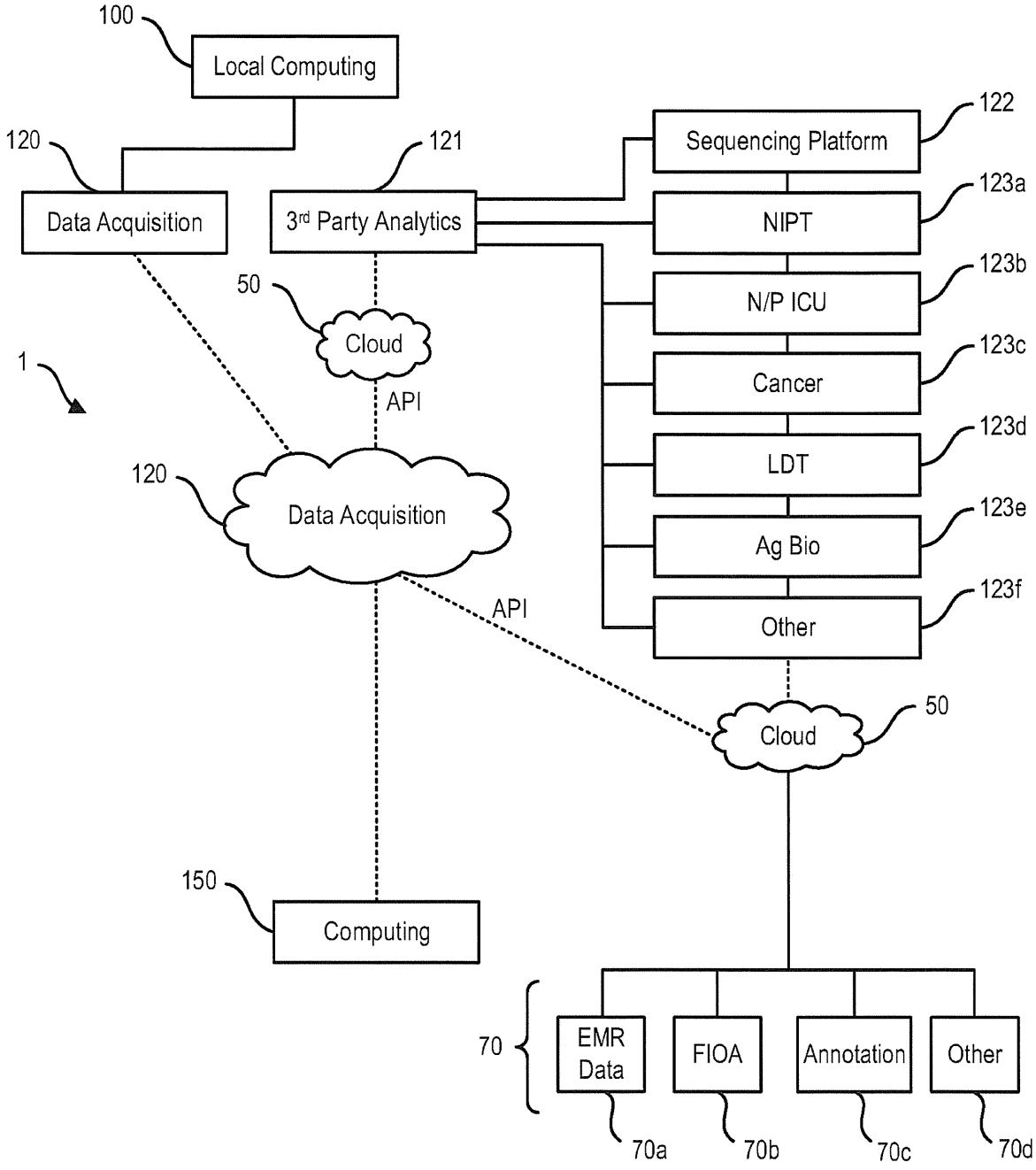

100
Local Computing
120
Data Acquisition
121
3rd Party Analytics
122
Sequencing Platform
123a
NIPT
123b
N/P ICU
123c
Cancer
123d
LDT
123e
Ag Bio
123f
Other
50
Cloud
API
1
120
Data Acquisition
API
50
Cloud
150
Computing
70
EMR Data
FIOA
Annotation
Other
70a
70b
70c
70d

FIG. 41C

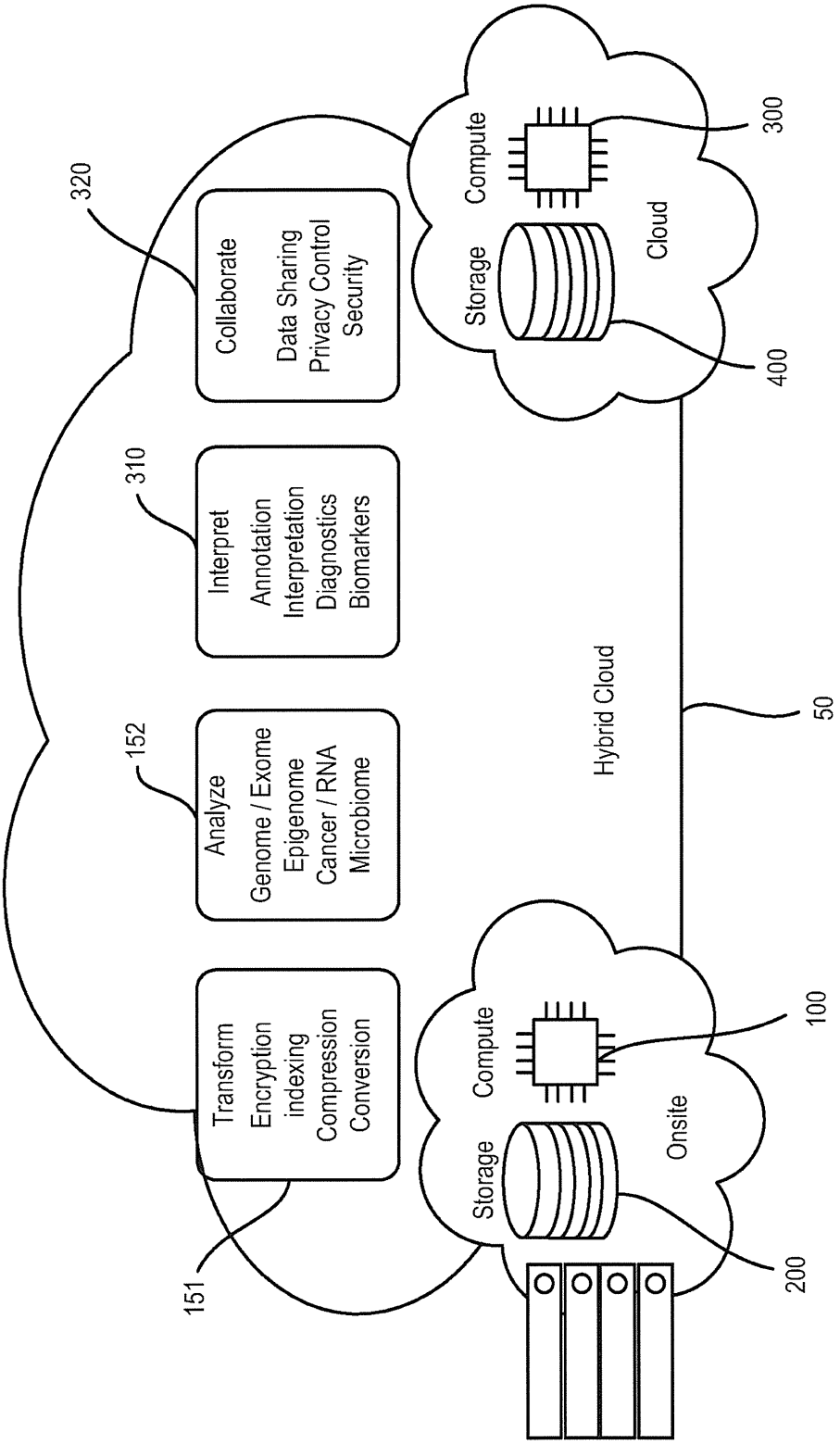

320
Collaborate
Data Sharing
Privacy Control
Security
310
Interpret
Annotation
Interpretation
Diagnostics
Biomarkers
152
Analyze
Genome / Exome
Epigenome
Cancer / RNA
Microbiome
Transform
Encryption
indexing
Compression
Conversion
151
Compute
Storage
Onsite
100
200
Hybrid Cloud
50
Compute
Storage
Cloud
300
400

FIG. 42A

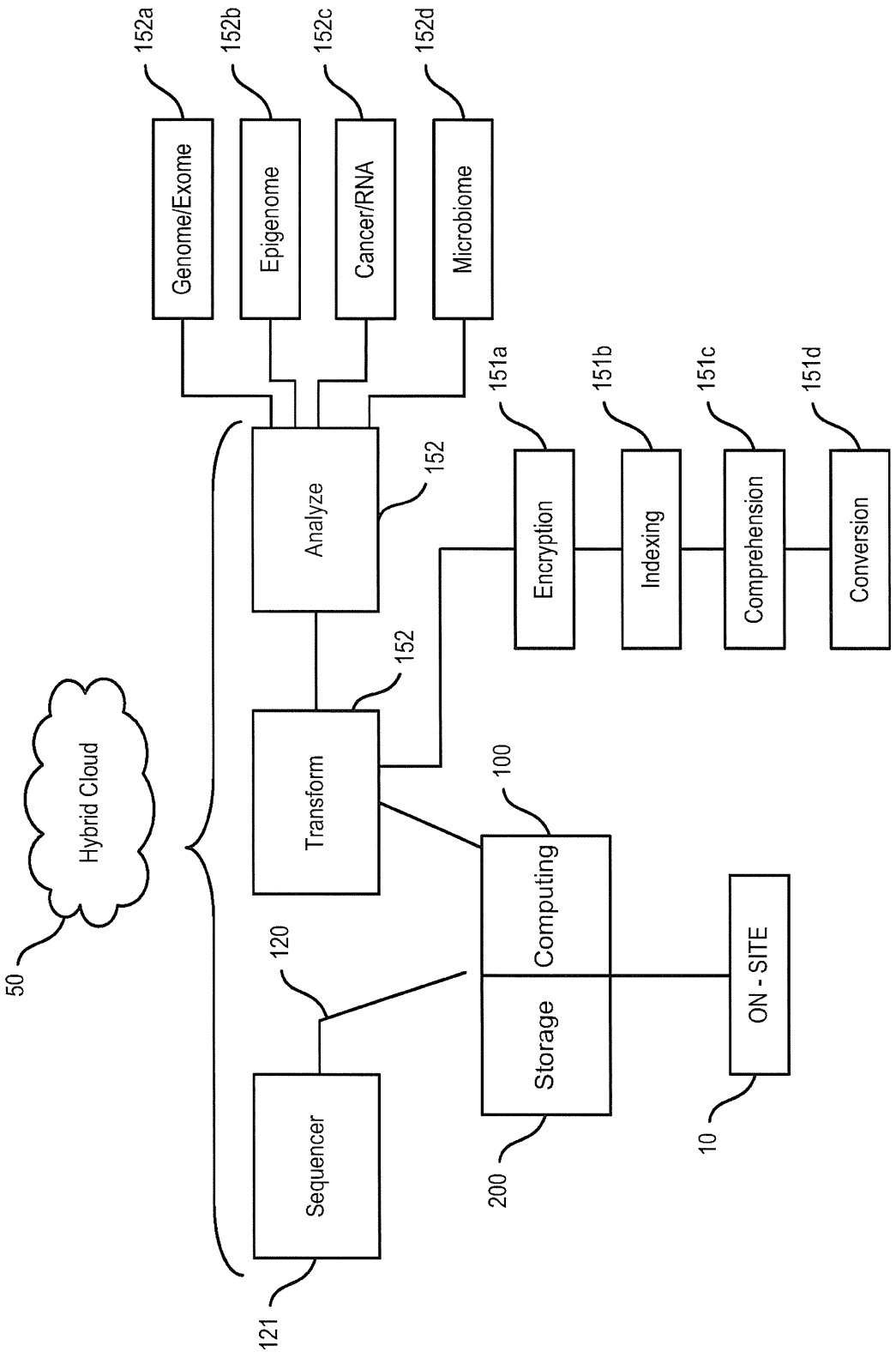

Hybrid Cloud
50
Sequencer
121
120
Transform
152
Analyze
152
Genome/Exome
Epigenome
Cancer/RNA
Microbiome
152a
152b
152c
152d
Encryption
Indexing
Comprehension
Conversion
151a
151b
151c
151d
Computing
Storage
100
200
ON - SITE
10

FIG. 42B

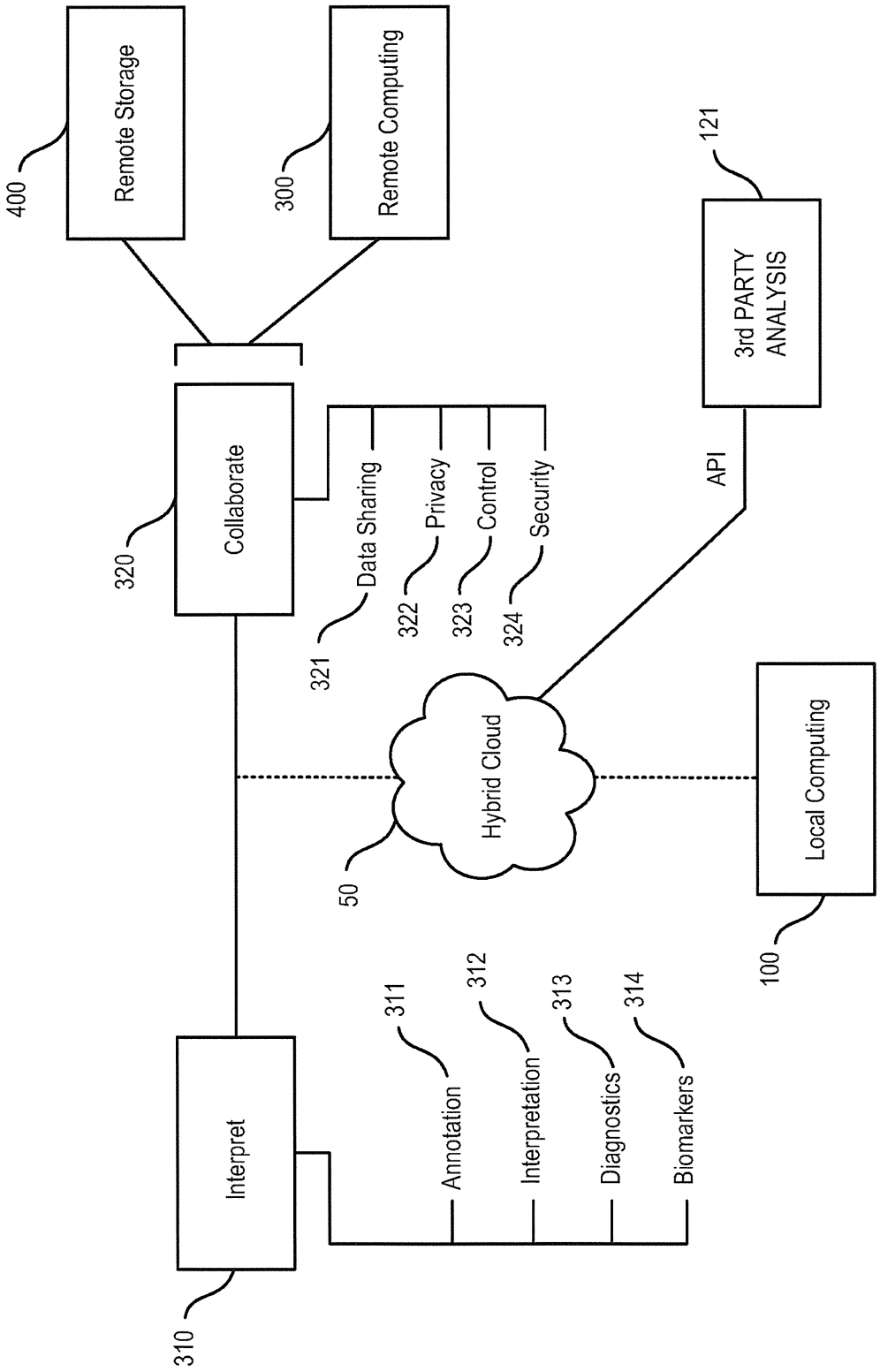

Remote Storage
400
Remote Computing
300
3rd PARTY ANALYSIS
121
Collaborate
320
Data Sharing
321
Privacy
322
Control
323
Security
324
API
Hybrid Cloud
50
Local Computing
100
Interpret
310
Annotation
311
Interpretation
312
Diagnostics
313
Biomarkers
314

FIG. 42C

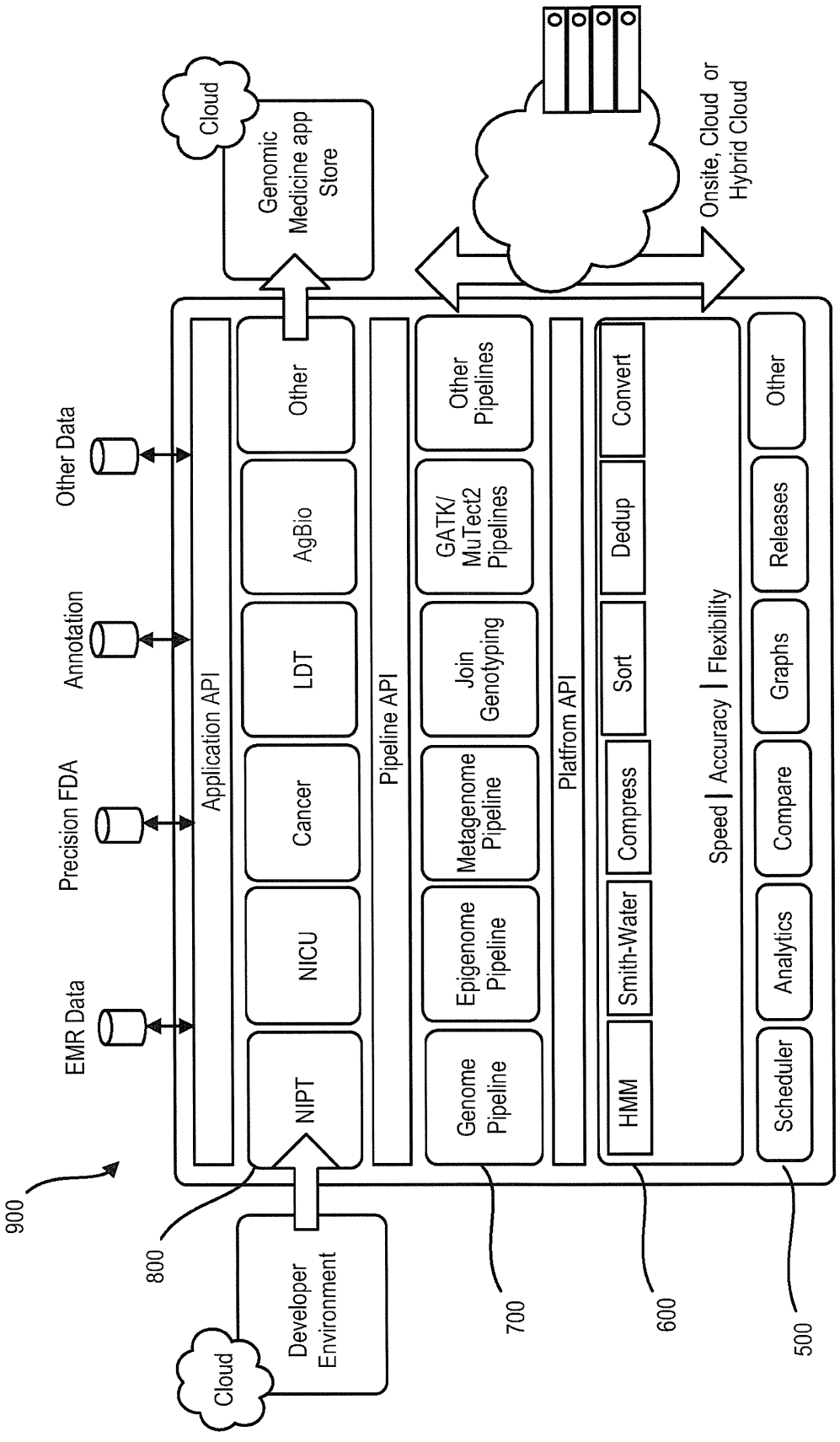

900
800
700
600
500
Cloud
Developer Environment
Cloud
Genomic Medicine app Store
Onsite, Cloud or Hybrid Cloud
EMR Data
Precision FDA
Annotation
Other Data
Application API
Pipeline API
Platfrom API
NIPT
NICU
Cancer
LDT
AgBio
Other
Genome Pipeline
Epigenome Pipeline
Metagenome Pipeline
Join Genotyping
GATK/ MuTect2 Pipelines
Other Pipelines
HMM
Smith-Water
Compress
Sort
Dedup
Convert
Speed | Accuracy | Flexibility
Scheduler
Analytics
Compare
Graphs
Releases
Other

FIG. 43A

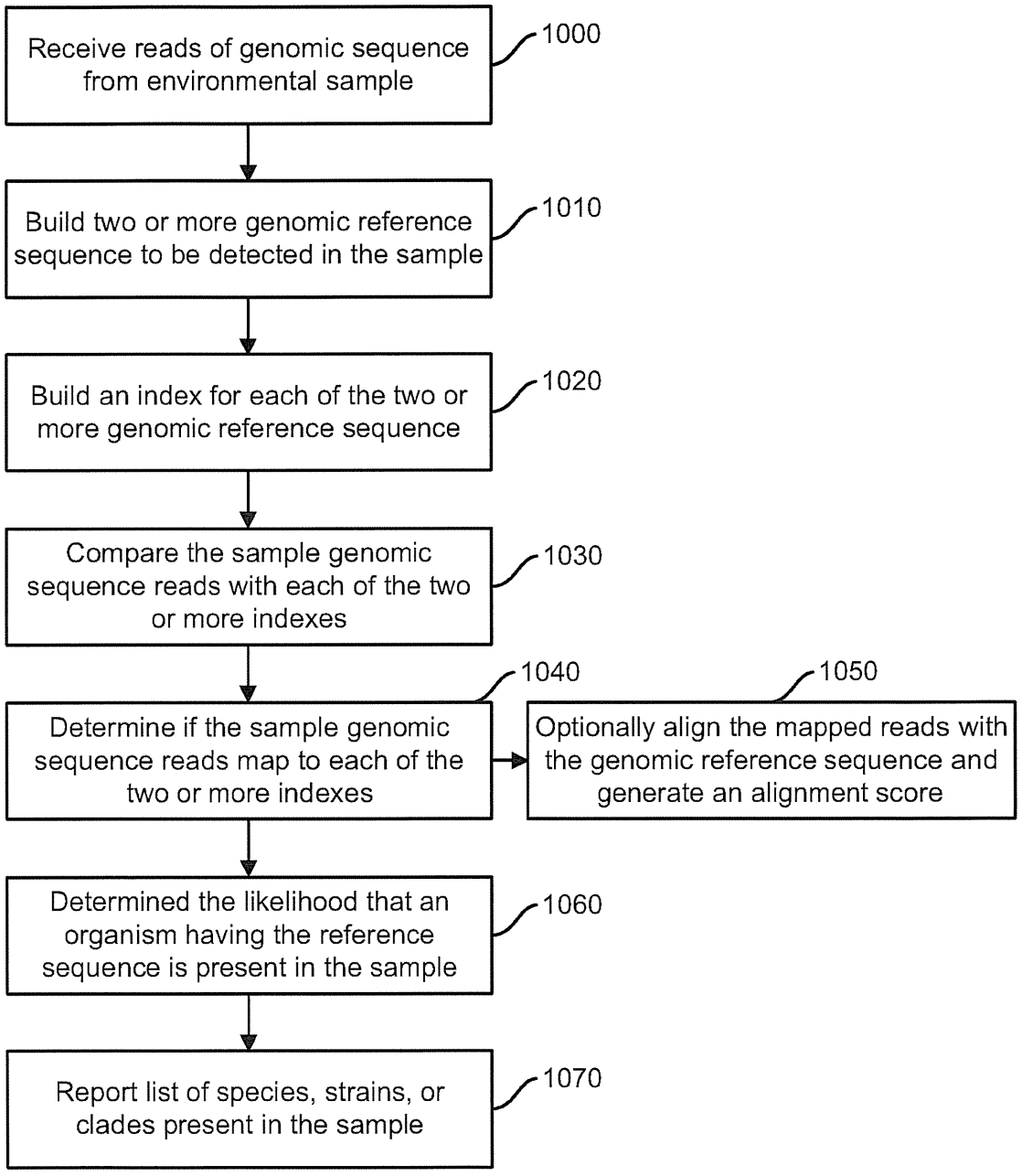

Receive reads of genomic sequence from environmental sample
1000
Build two or more genomic reference sequence to be detected in the sample
1010
Build an index for each of the two or more genomic reference sequence
1020
Compare the sample genomic sequence reads with each of the two or more indexes
1030
Determine if the sample genomic sequence reads map to each of the two or more indexes
1040
Optionally align the mapped reads with the genomic reference sequence and generate an alignment score
1050
Determined the likelihood that an organism having the reference sequence is present in the sample
1060
Report list of species, strains, or clades present in the sample
1070

FIG. 43B

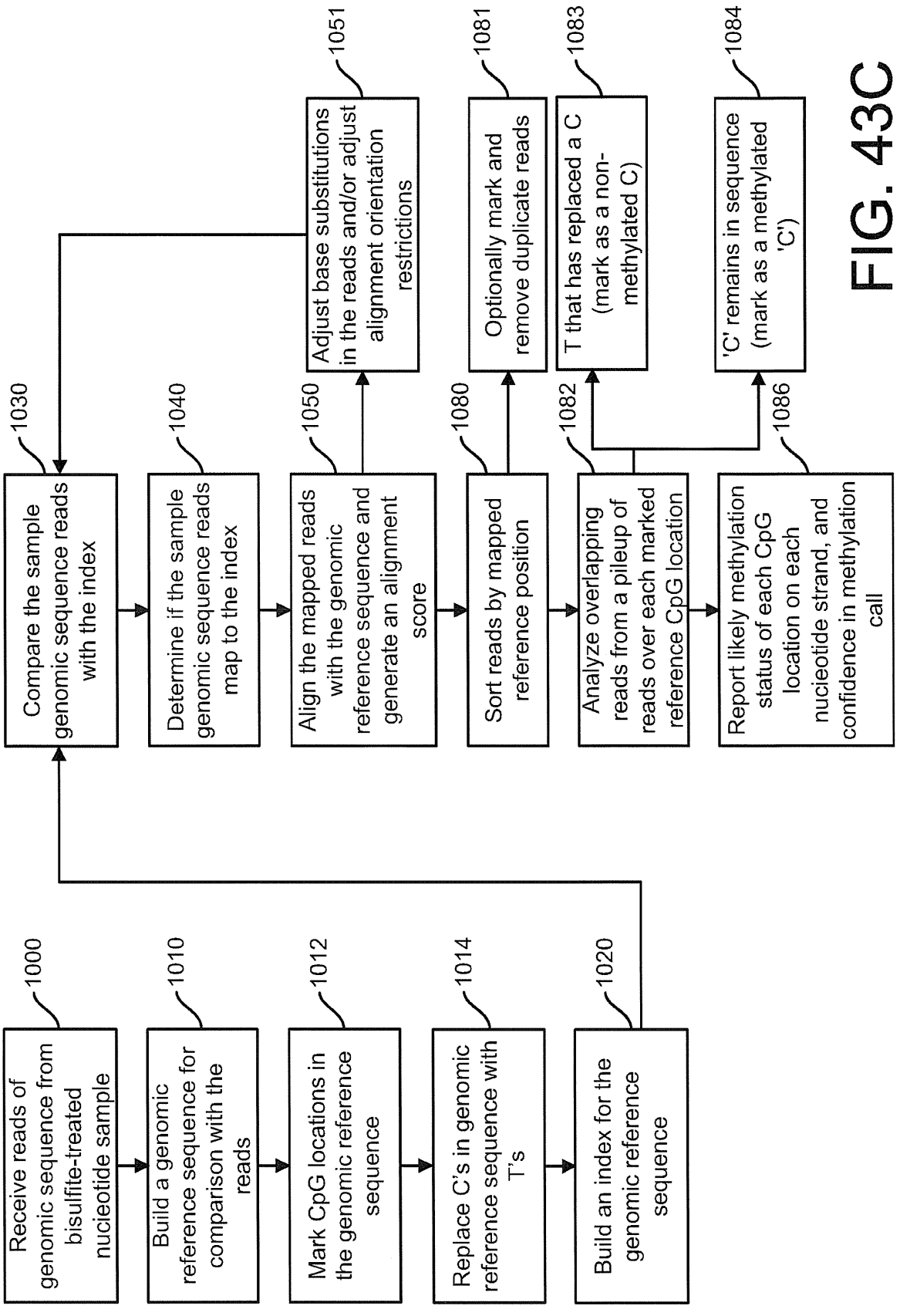

Receive reads of genomic sequence from bisulfite-treated nucieotide sample
1000
Build a genomic reference sequence for comparison with the reads
1010
Mark CpG locations in the genomic reference sequence
1012
Replace C's in genomic reference sequence with T's
1014
Build an index for the genomic reference sequence
1020
Compare the sample genomic sequence reads with the index
1030
Determine if the sample genomic sequence reads map to the index
1040
Align the mapped reads with the genomic reference sequence and generate an alignment score
1050
Adjust base substitutions in the reads and/or adjust alignment orientation restrictions
1051
Sort reads by mapped reference position
1080
Optionally mark and remove duplicate reads
1081
Analyze overlapping reads from a pileup of reads over each marked reference CpG location
1082
T that has replaced a C (mark as a non-methylated C)
1083
Report likely methylation status of each CpG location on each nucieotide strand, and confidence in methylation call
1086
'C' remains in sequence (mark as a methylated 'C')
1084

FIG. 43C

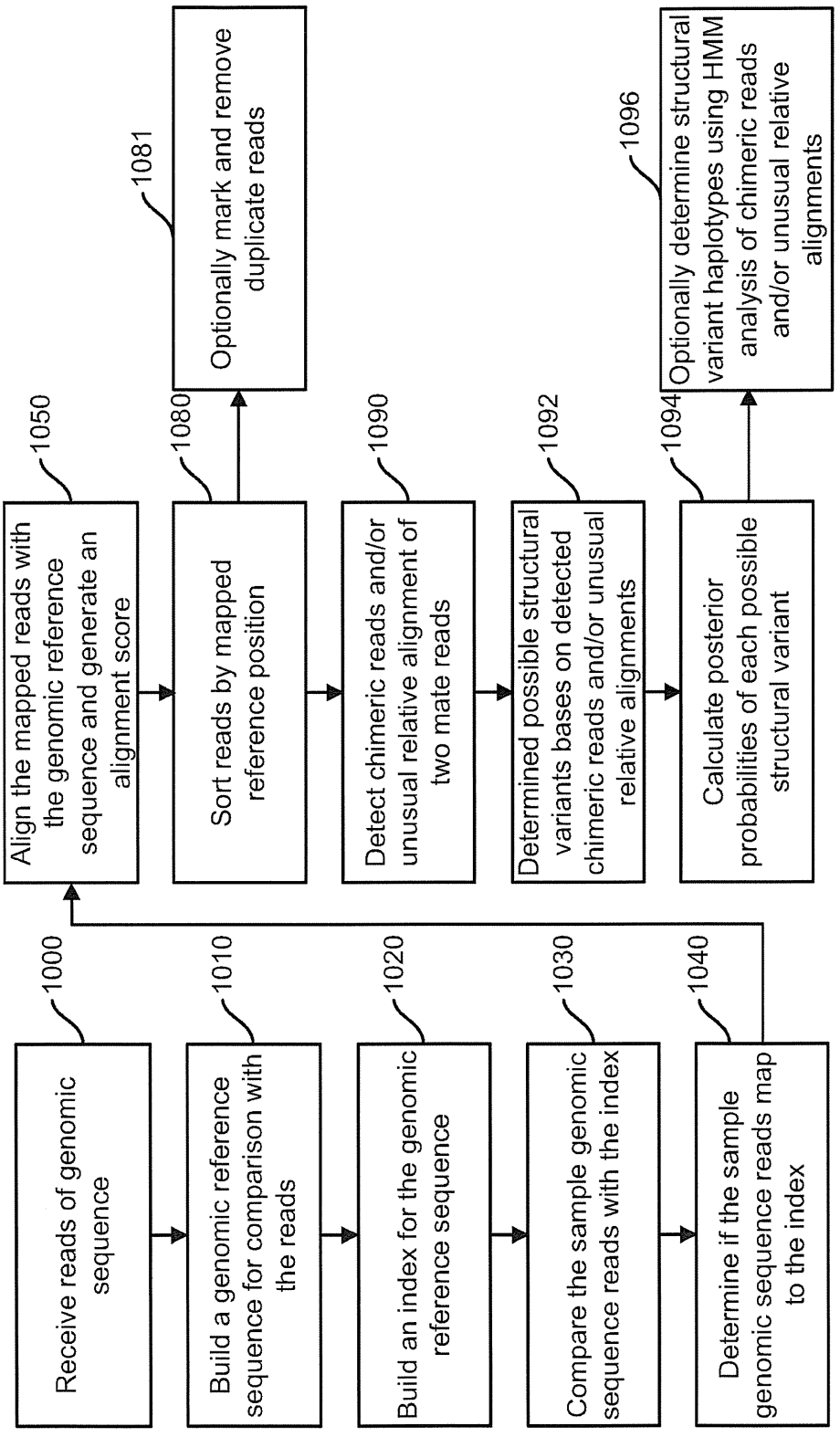

Receive reads of genomic sequence
1000
Build a genomic reference sequence for comparison with the reads
1010
Build an index for the genomic reference sequence
1020
Compare the sample genomic sequence reads with the index
1030
Determine if the sample genomic sequence reads map to the index
1040
Align the mapped reads with the genomic reference sequence and generate an alignment score
1050
Sort reads by mapped reference position
1080
Optionally mark and remove duplicate reads
1081
Detect chimeric reads and/or unusual relative alignment of two mate reads
1090
Determined possible structural variants bases on detected chimeric reads and/or unusual relative alignments
1092
Calculate posterior probabilities of each possible structural variant
1094
Optionally determine structural variant haplotypes using HMM analysis of chimeric reads and/or unusual relative alignments
1096

FIG. 43D

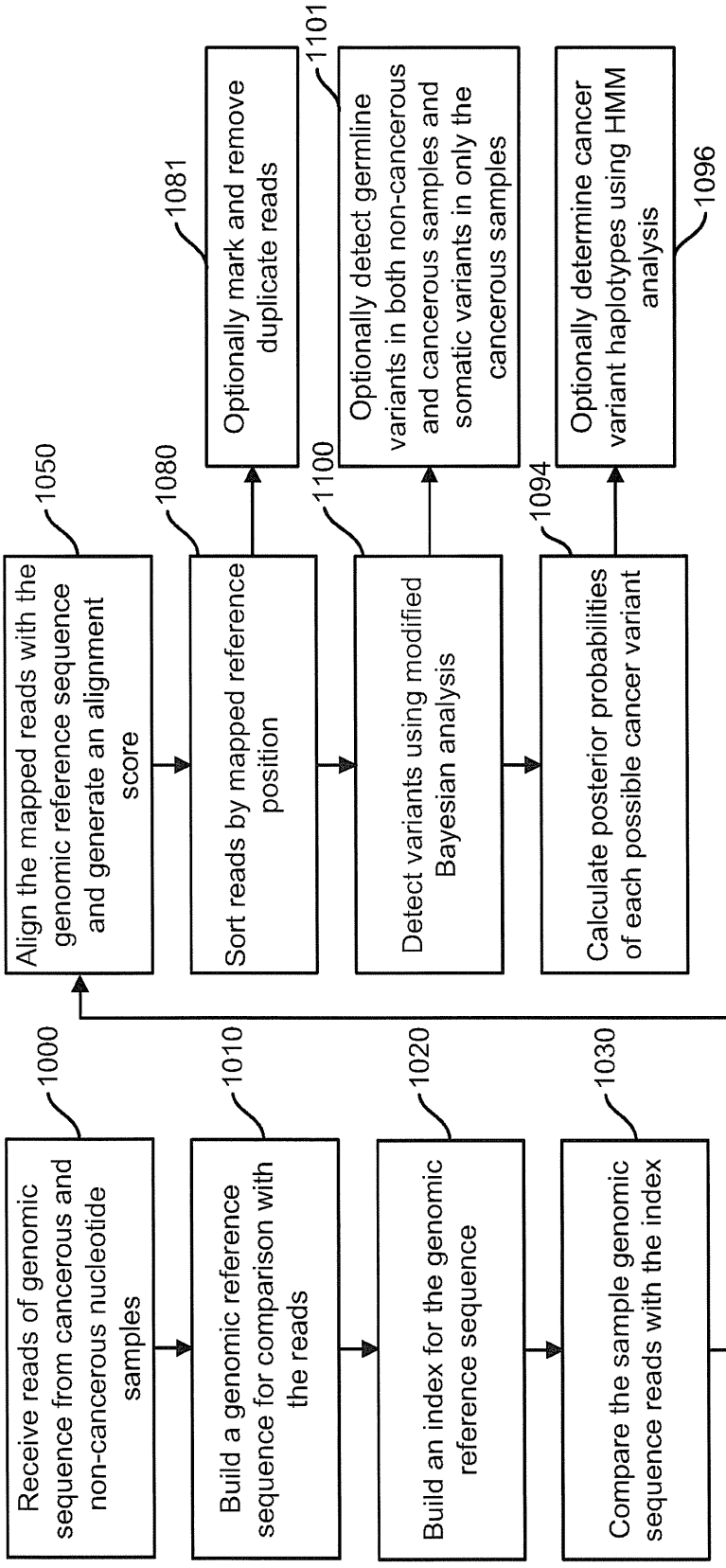

Receive reads of genomic sequence from cancerous and non-cancerous nucleotide samples
1000
Build a genomic reference sequence for comparison with the reads
1010
Build an index for the genomic reference sequence
1020
Compare the sample genomic sequence reads with the index
1030
Determined if the sample genomic sequence reads map to the index
1040
Align the mapped reads with the genomic reference sequence and generate an alignment score
1050
Sort reads by mapped reference position
1080
Optionally mark and remove duplicate reads
1081
Detect variants using modified Bayesian analysis
1100
Optionally detect germline variants in both non-cancerous and cancerous samples and somatic variants in only the cancerous samples
1101
Calculate posterior probabilities of each possible cancer variant
1094
Optionally determine cancer variant haplotypes using HMM analysis
1096

FIG. 43E

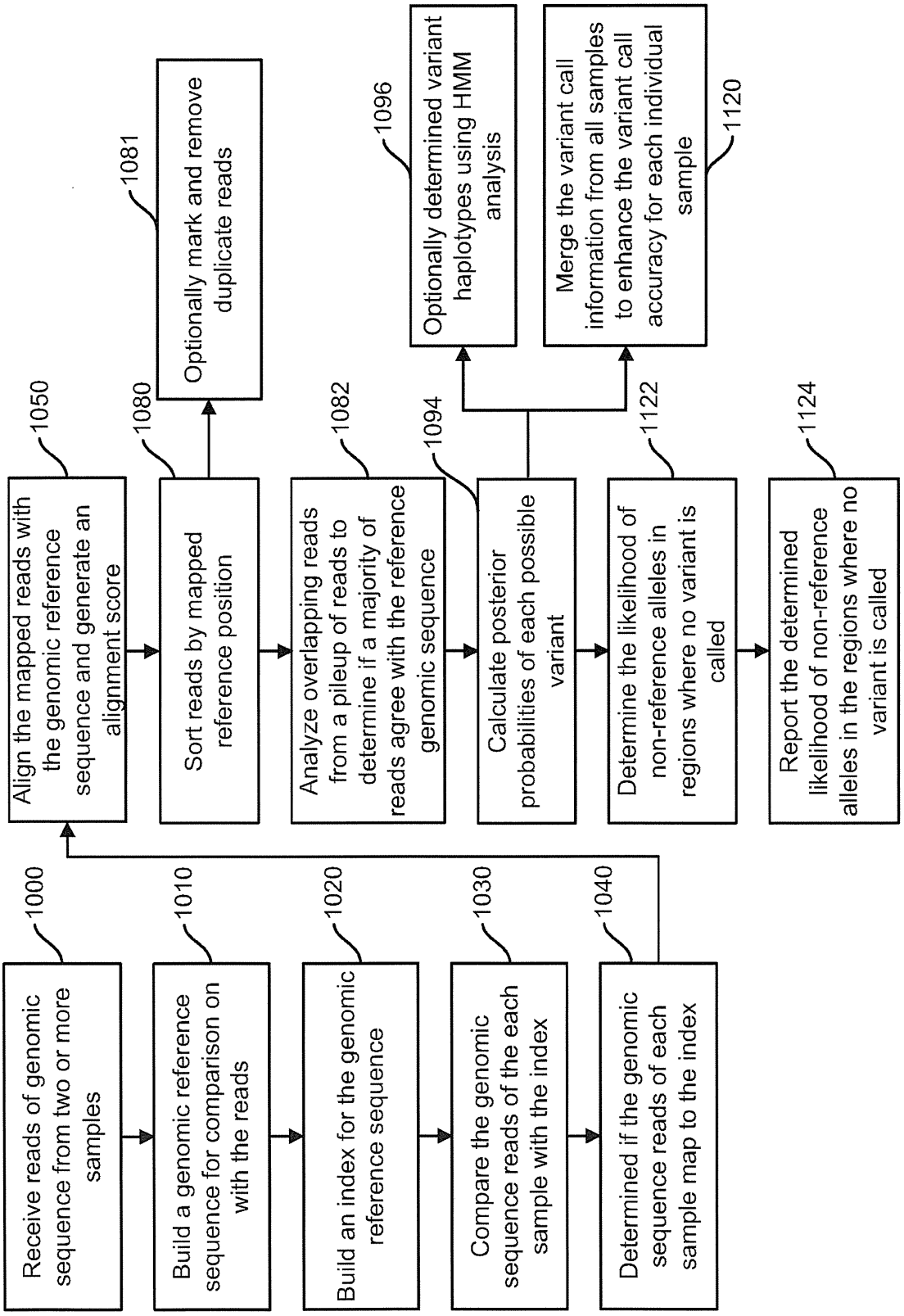

Receive reads of genomic sequence from two or more samples
1000
Build a genomic reference sequence for comparison on with the reads
1010
Build an index for the genomic reference sequence
1020
Compare the genomic sequence reads of the each sample with the index
1030
Determined if the genomic sequence reads of each sample map to the index
1040
Align the mapped reads with the genomic reference sequence and generate an alignment score
1050
Sort reads by mapped reference position
1080
Optionally mark and remove duplicate reads
1081
Analyze overlapping reads from a pileup of reads to determine if a majority of reads agree with the reference genomic sequence
1082
Calculate posterior probabilities of each possible variant
1094
Optionally determined variant haplotypes using HMM analysis
1096
Determine the likelihood of non-reference alleles in regions where no variant is called
1122
Merge the variant call information from all samples to enhance the variant call accuracy for each individual sample
1120
Report the determined likelihood of non-reference alleles in the regions where no variant is called
1124

FIG. 43F

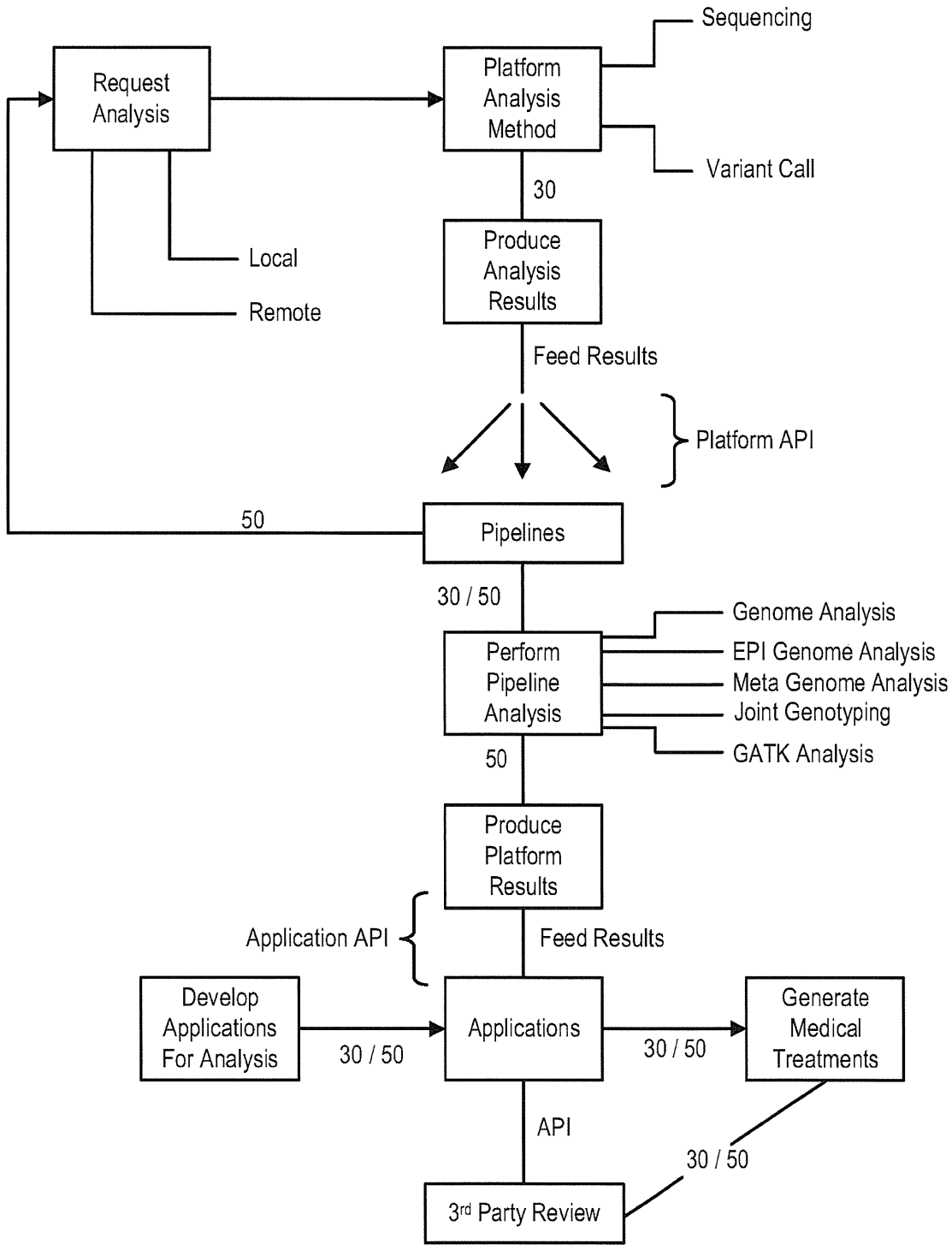

Request Analysis
Platform Analysis Method
Sequencing
Variant Call
30
Local
Remote
Produce Analysis Results
Feed Results
Platform API
50
Pipelines
30 / 50
Genome Analysis
EPI Genome Analysis
Meta Genome Analysis
Joint Genotyping
GATK Analysis
Perform Pipeline Analysis
50
Produce Platform Results
Application API
Feed Results
Develop Applications For Analysis
30 / 50
Applications
30 / 50
Generate Medical Treatments
API
30 / 50
3rd Party Review

FIG. 44

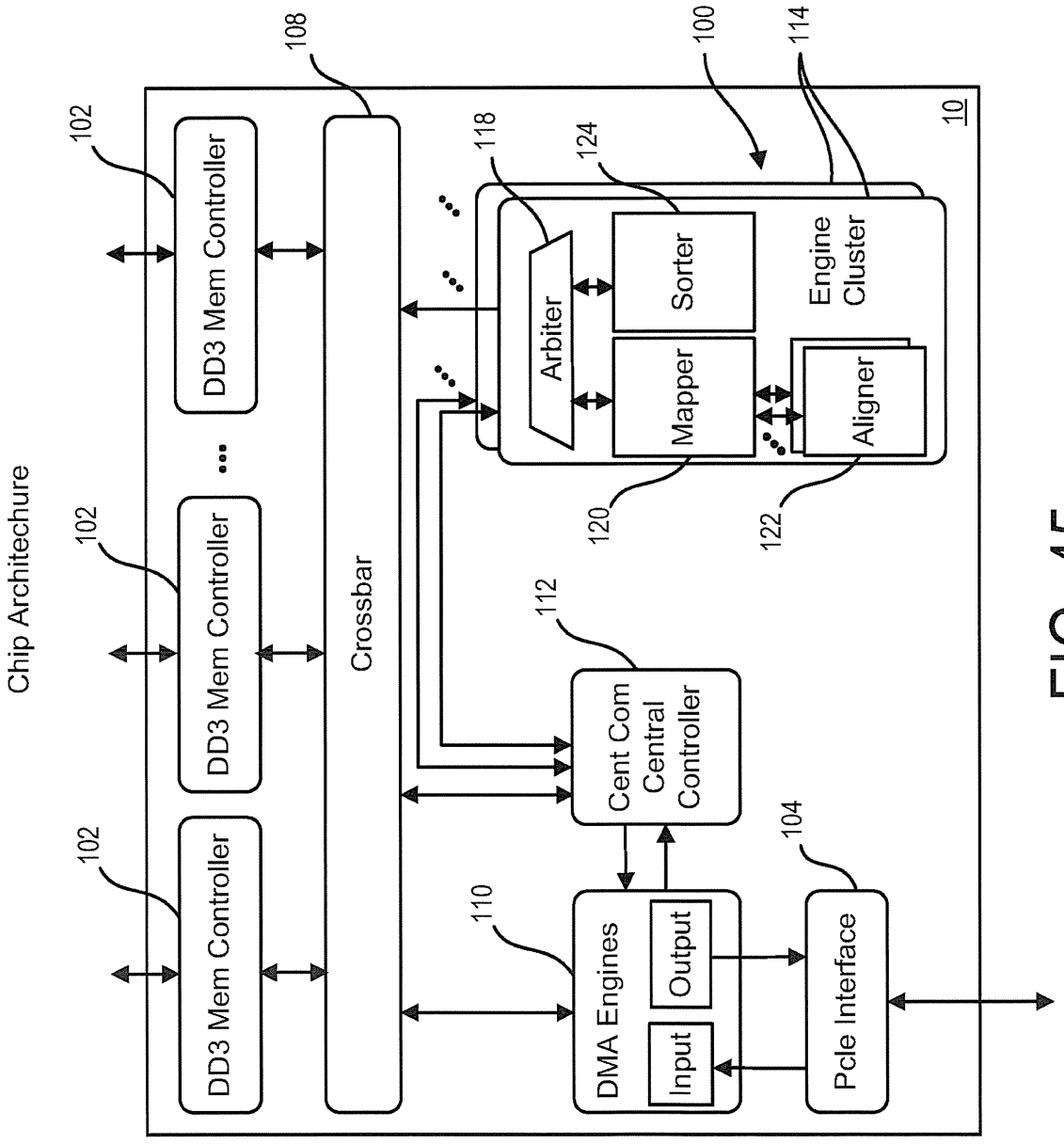

Chip Architechure
DD3 Mem Controller
DD3 Mem Controller
DD3 Mem Controller
Crossbar
Arbiter
Sorter
Mapper
Aligner
Engine Cluster
Cent Com Central Controller
DMA Engines
Input
Output
Pcle Interface
102
102
102
108
100
114
118
124
120
122
112
110
104
10

FIG. 45

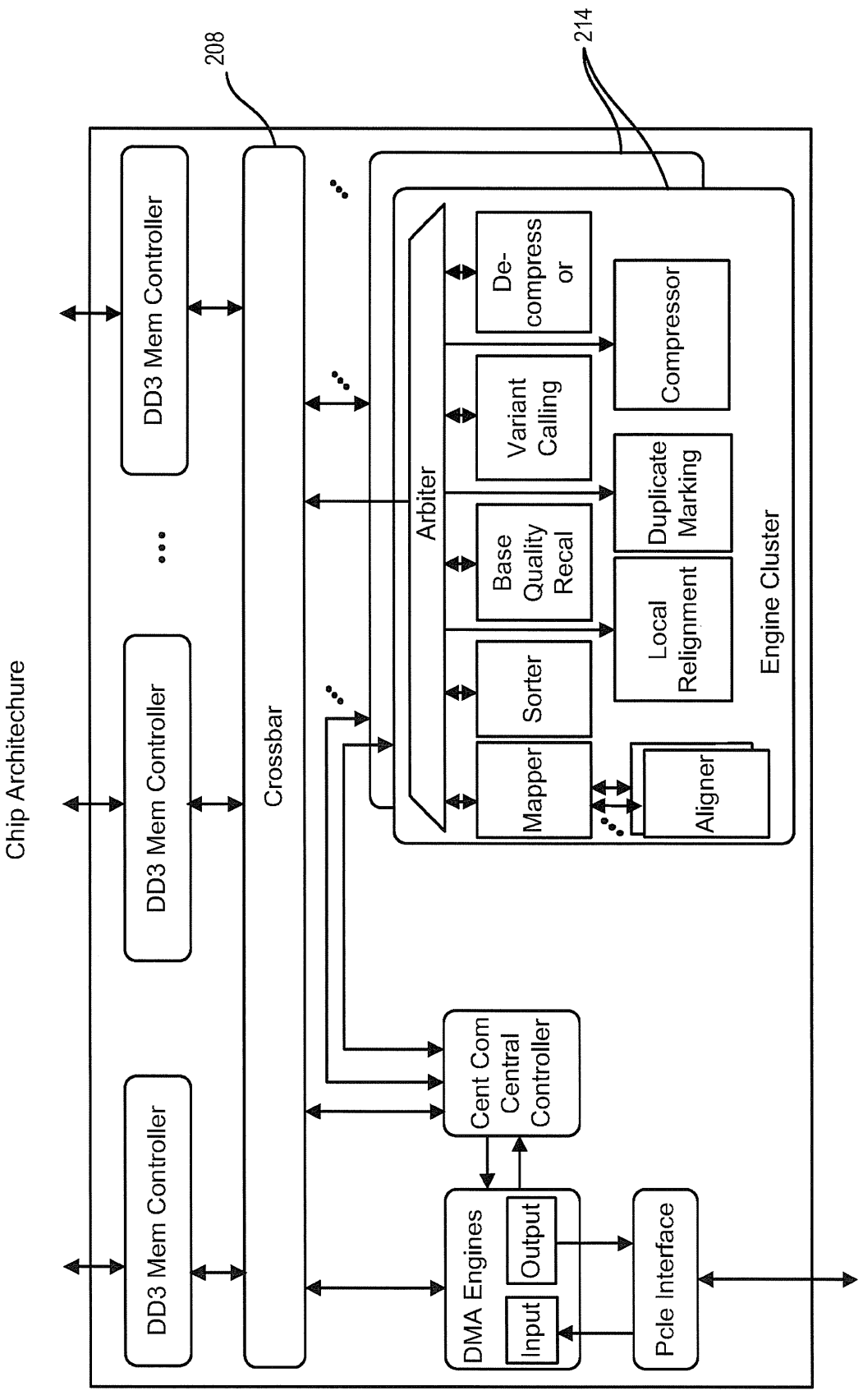

Chip Architechure
208
214
DD3 Mem Controller
DD3 Mem Controller
DD3 Mem Controller
Crossbar
Arbiter
De-compressor
Variant Calling
Base Quality Recal
Sorter
Mapper
Compressor
Duplicate Marking
Local Relignment
Aligner
Engine Cluster
Cent Com Central Controller
DMA Engines
Input
Output
Pcle Interface

FIG. 46

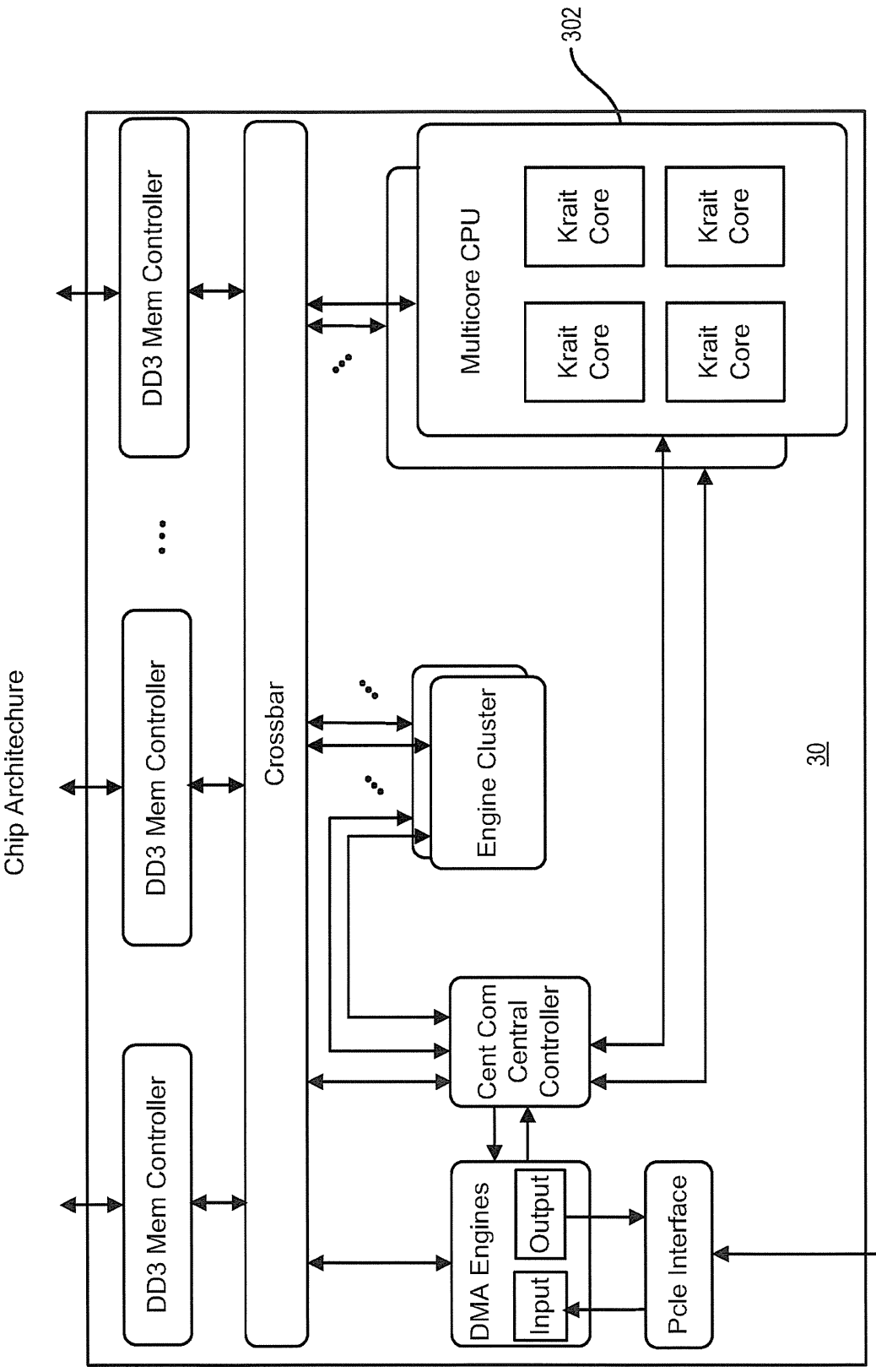

Chip Architechure
DD3 Mem Controller
DD3 Mem Controller
DD3 Mem Controller
Crossbar
Multicore CPU
Krait Core
Krait Core
Krait Core
Krait Core
302
Engine Cluster
Cent Com Central Controller
DMA Engines
Input
Output
Pcie Interface
30

FIG. 47

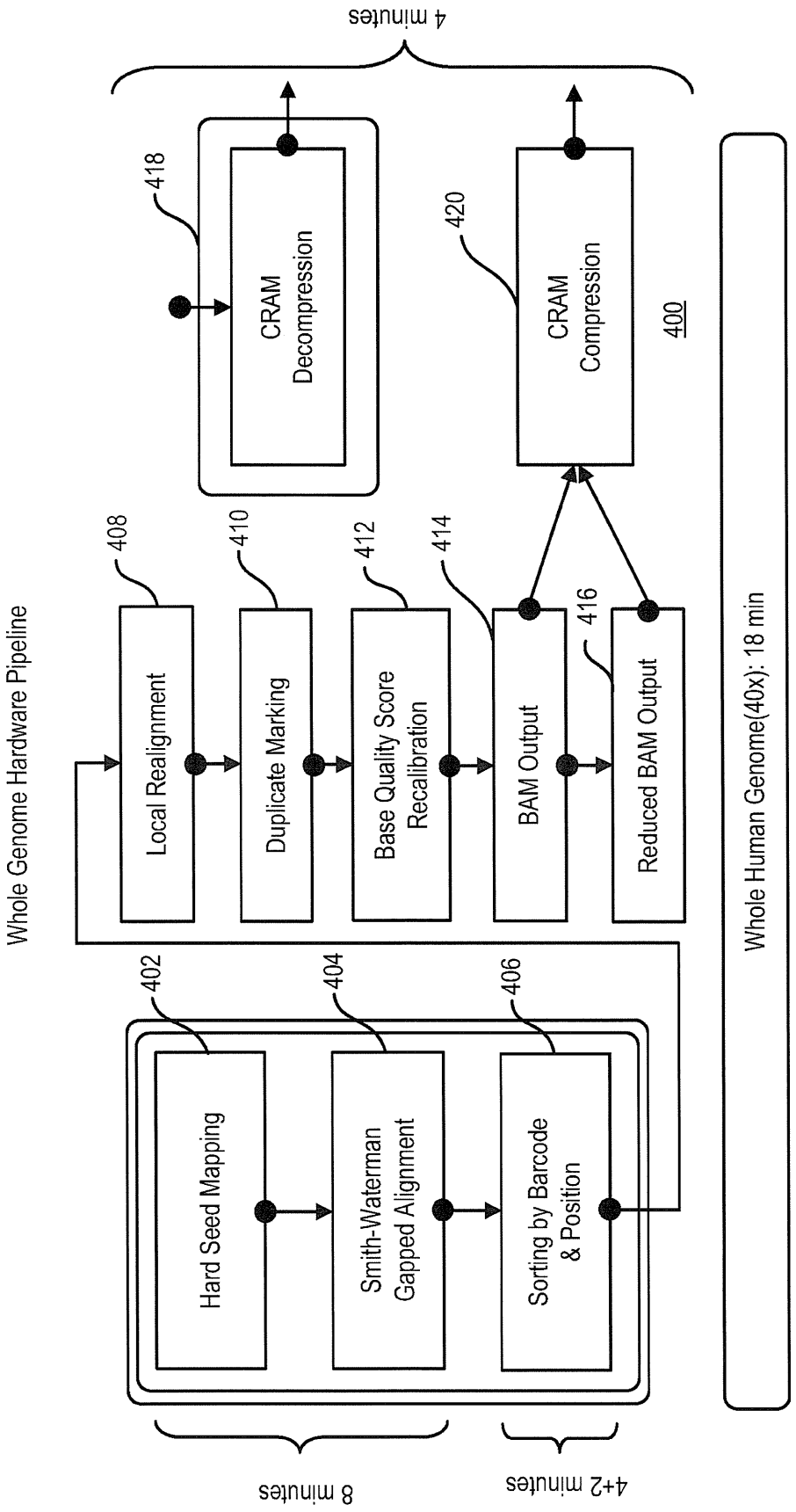

Whole Genome Hardware Pipeline
Hard Seed Mapping
Smith-Waterman Gapped Alignment
Sorting by Barcode & Position
Local Realignment
Duplicate Marking
Base Quality Score Recalibration
BAM Output
Reduced BAM Output
CRAM Decompression
CRAM Compression
402
404
406
408
410
412
414
416
418
420
400
8 minutes
4+2 minutes
4 minutes
Whole Human Genome(40x): 18 min

FIG. 48

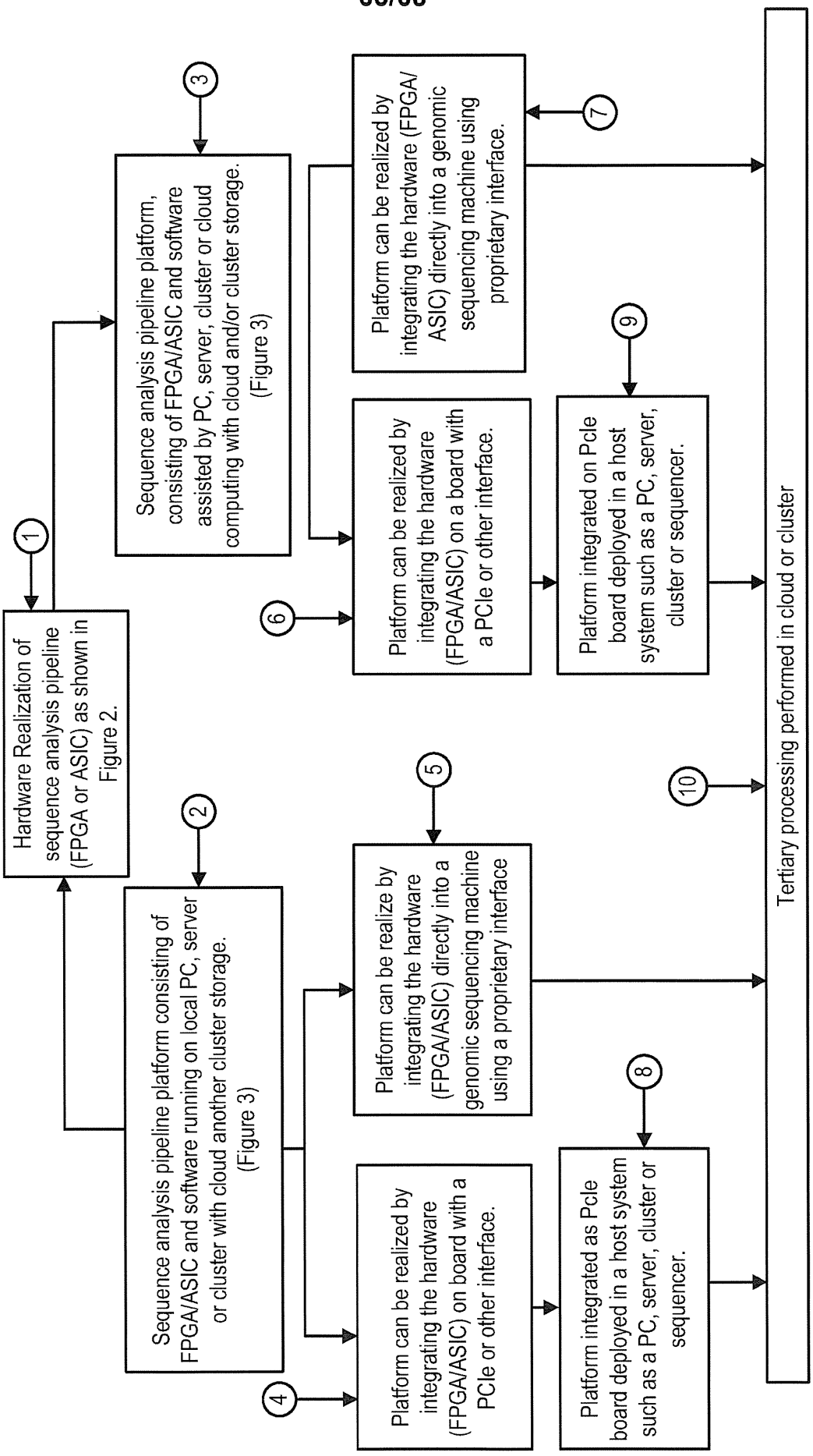

Hardware Realization of sequence analysis pipeline (FPGA or ASIC) as shown in Figure 2.
1
Sequence analysis pipeline platform consisting of FPGA/ASIC and software running on local PC, server or cluster with cloud another cluster storage. (Figure 3)
2
Sequence analysis pipeline platform, consisting of FPGA/ASIC and software assisted by PC, server, cluster or cloud computing with cloud and/or cluster storage. (Figure 3)
3
Platform can be realized by integrating the hardware (FPGA/ASIC) on board with a PCIe or other interface.
4
Platform can be realize by integrating the hardware (FPGA/ASIC) directly into a genomic sequencing machine using a proprietary interface
5
Platform can be realized by integrating the hardware (FPGA/ASIC) on a board with a PCIe or other interface.
6
Platform can be realized by integrating the hardware (FPGA/ASIC) directly into a genomic sequencing machine using proprietary interface.
7
Platform integrated as Pcle board deployed in a host system such as a PC, server, cluster or sequencer.
8
Platform integrated on Pcle board deployed in a host system such as a PC, server, cluster or sequencer.
9
10
Tertiary processing performed in cloud or cluster

FIG. 49

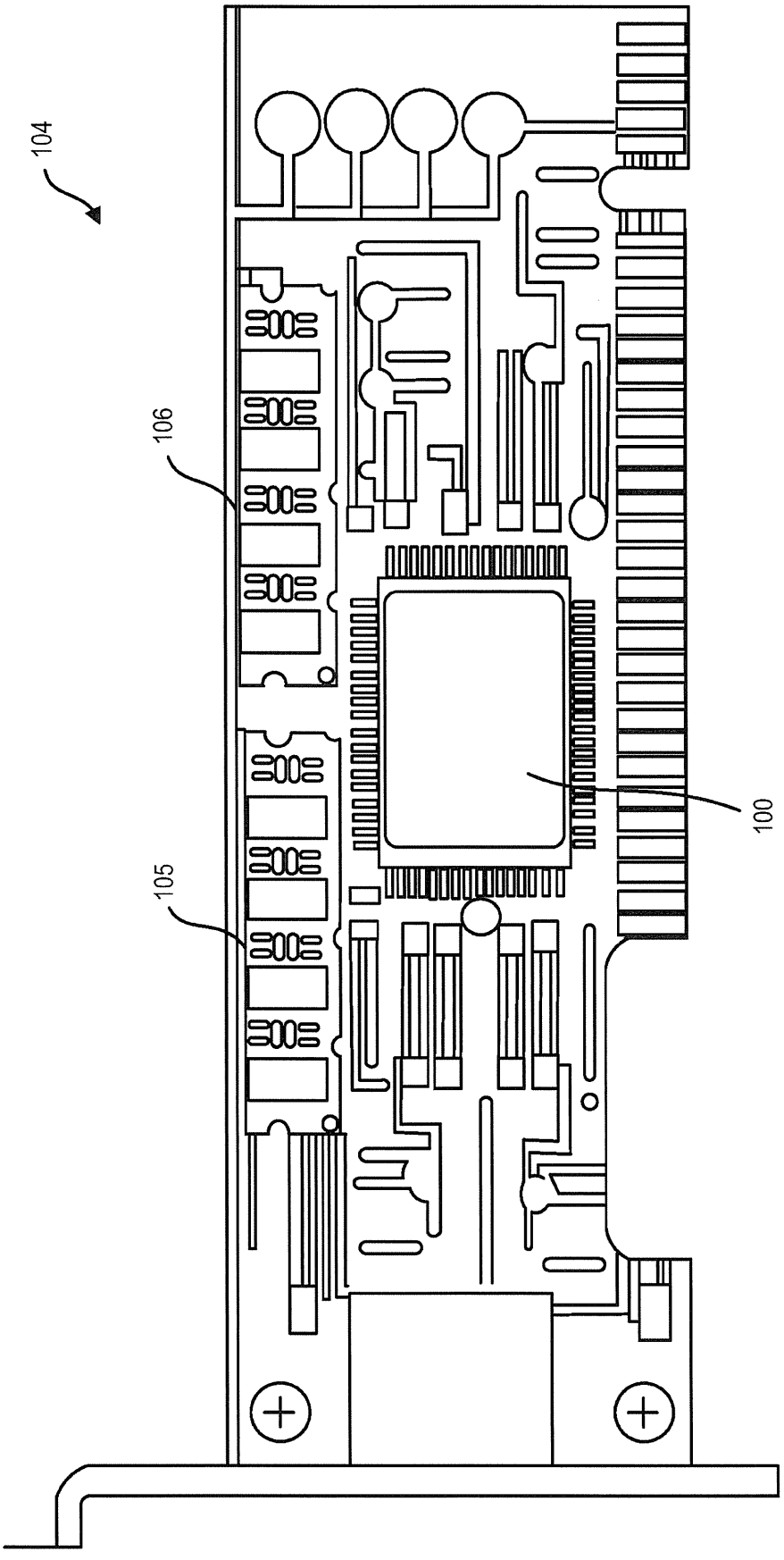


FIG. 50A

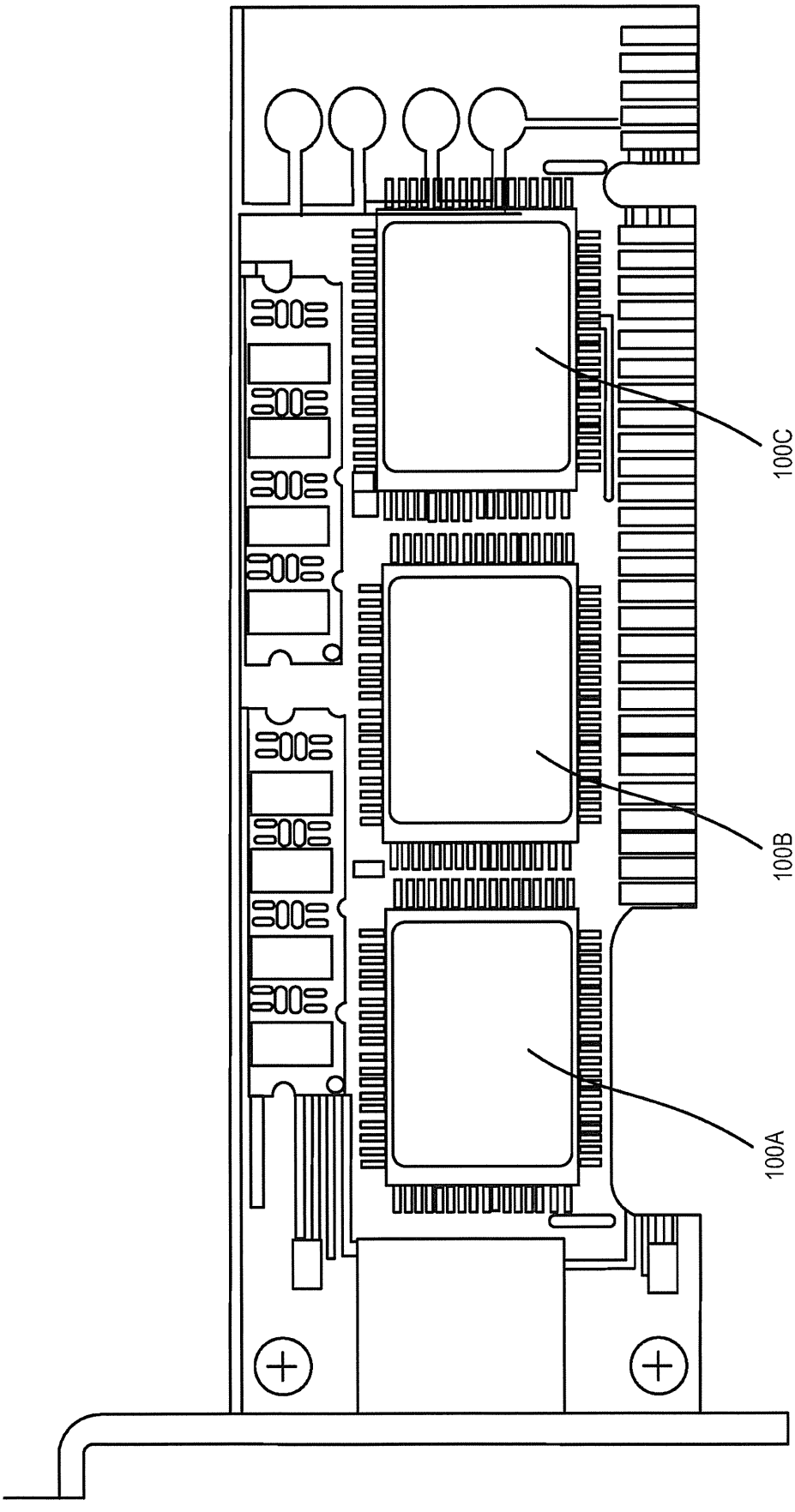


FIG. 50B

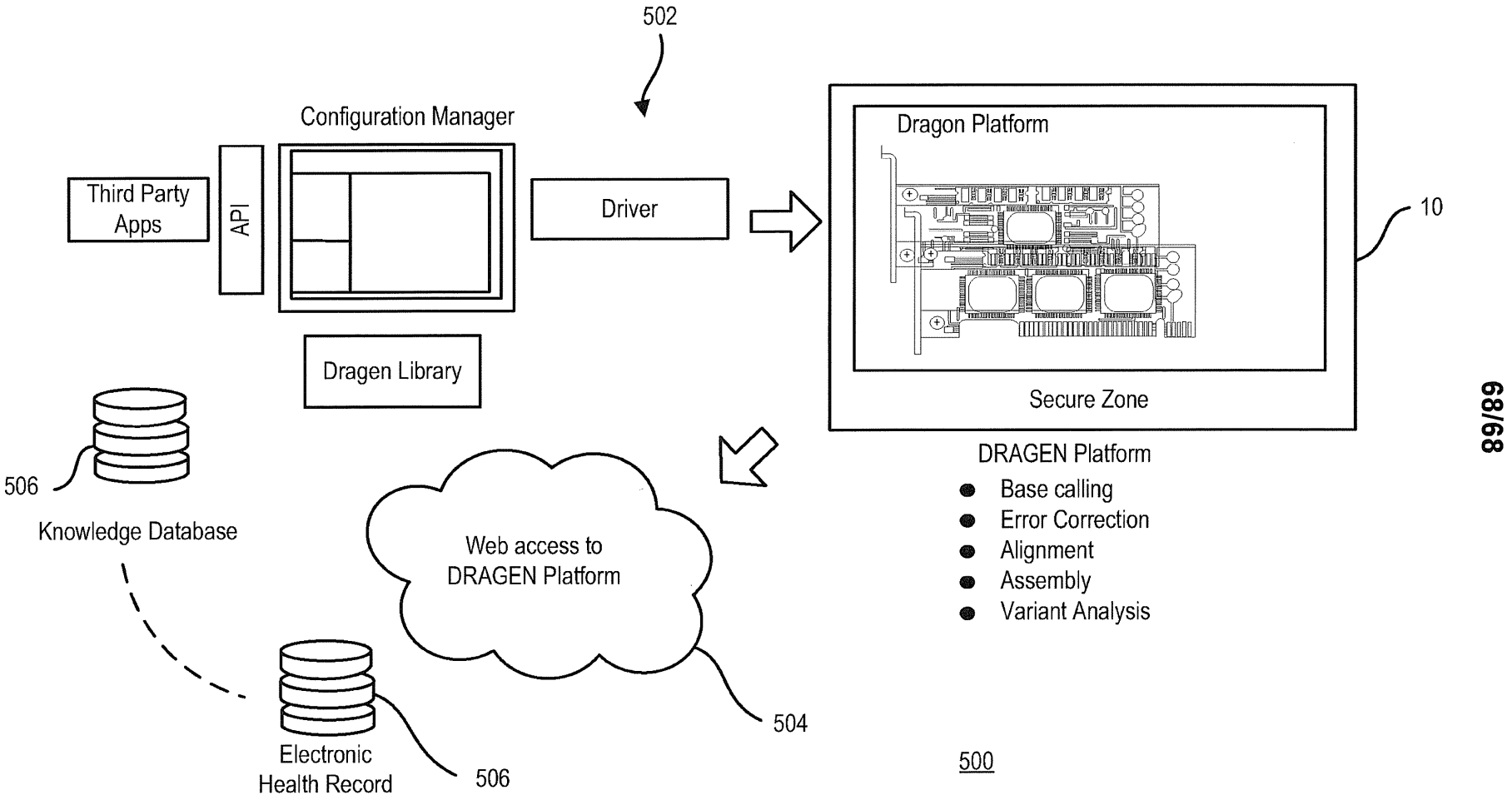

502
Configuration Manager
Third Party Apps
API
Driver
Dragon Platform
10
Dragen Library
Secure Zone
506
Knowledge Database
DRAGEN Platform
Base calling
Error Correction
Alignment
Assembly
Variant Analysis
Web access to DRAGEN Platform
504
Electronic Health Record
506
500

FIG. 51