RSF conversion to/from SEG-Y and ASCII

Prof. Daniel Leal Macedo

Segundo Curso de Inverno do Observatório Sismológico - UnB

6 e 7 de Junho de 2017

SEG-Y

The SEG-Y format is one of several tape standards developed by the Society of Exploration Geophysicists (SEG). It is the most common format used for seismic data in the exploration and production industry. However, it was created in 1973 and many different 'modernized' flavors exist.

SEG-Y

SEG-Y was designed for storing a single line of seismic data on IBM 9-track tapes attached to IBM mainframe computers. Most of the variations in modern SEG-Y varieties result from trying to overcome these limitations.

Goals

1. SEG-Y components
2. Basics of acquisition geometry
3. Conversion

SEG-Y components

The official standard SEG-Y consists of the following components:

- a 3200-byte EBCDIC (Extended Binary Coded Decimal Interchange Code) descriptive reel header records
- a 400-byte binary reel header record
- trace records consisting of
  - a 240-byte binary trace header
  - trace data

SEG-Y components

The official standard SEG-Y consists of the following components:

- a 3200-byte EBCDIC (Extended Binary Coded Decimal Interchange Code) descriptive reel header records
- a 400-byte binary reel header record
- trace records consisting of
  - a 240-byte binary trace header
  - trace data

SEG-Y components

The official standard SEG-Y consists of the following components:

- a 3200-byte EBCDIC (Extended Binary Coded Decimal Interchange Code) descriptive reel header records
- a 400-byte binary reel header record
- trace records consisting of
  - a 240-byte binary trace header
  - trace data

SEG-Y components

The official standard SEG-Y consists of the following components:

- a 3200-byte EBCDIC (Extended Binary Coded Decimal Interchange Code) descriptive reel header records
- a 400-byte binary reel header record
- trace records consisting of
  - a 240-byte binary trace header
  - trace data

SEG-Y components

The official standard SEG-Y consists of the following components:

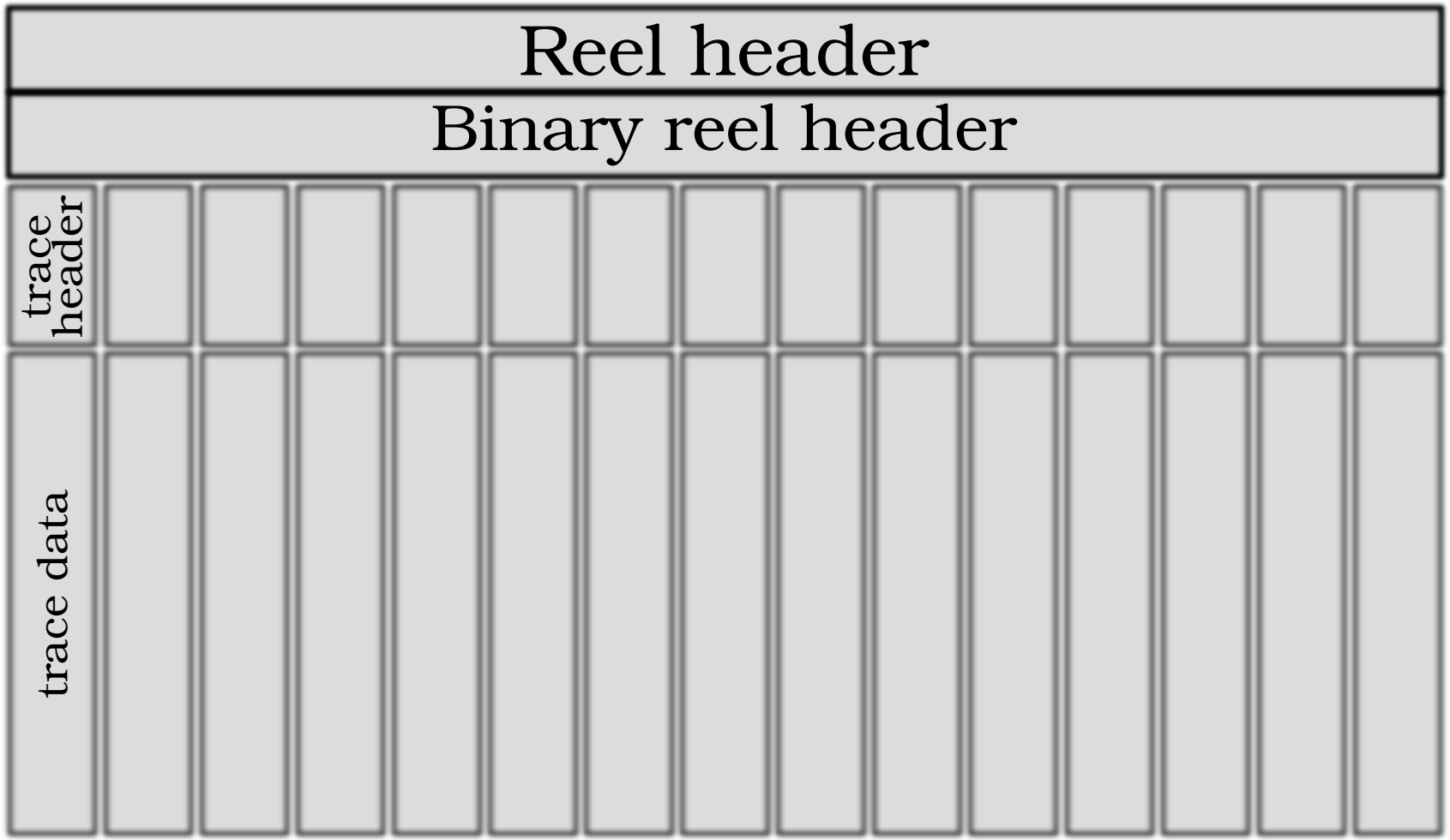
- a 3200-byte EBCDIC (Extended Binary Coded Decimal Interchange Code) descriptive reel header records
- a 400-byte binary reel header record
- trace records consisting of
  - a 240-byte binary trace header
  - trace data

SEG-Y components

The official standard SEG-Y consists of the following components:

- a 3200-byte EBCDIC (Extended Binary Coded Decimal Interchange Code) descriptive reel header records
- a 400-byte binary reel header record
- trace records consisting of
  - a 240-byte binary trace header
  - trace data

SEG-Y components



The SEG-Y EBCDIC Reel Header

The EBCDIC reel header is equivalent to 40 IBM punch-cards (EBCDIC? punchcards? Welcome to the 70's, man!). The official layout of these 80-character cards is the EBCDIC equivalent of the following:

```
1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
C 1 CLIENT                          COMPANY                              CREW NO
C 2 LINE             AREA                              MAP ID
C 3 REEL NO            DAY-START OF REEL     YEAR      OBSERVER
C 4 INSTRUMENT: MFG             MODEL             SERIAL NO
C 5 DATA TRACES/RECORD         AUXILIARY TRACES/RECORD          CDP FOLD
C 6 SAMPLE INTERVAL          SAMPLES/TRACE       BITS/IN      BYTES/SAMPLE
C 7 RECORDING FORMAT         FORMAT THIS REEL          MEASUREMENT SYSTEM
C 8 SAMPLE CODE: FLOATING PT     FIXED PT     FIXED PT-GAIN     CORRELATED
C 9 GAIN  TYPE: FIXED     BINARY     FLOATING POINT     OTHER
C10 FILTERS: ALIAS     HZ  NOTCH     HZ  BAND     -     HZ  SLOPE    -    DB/OCT
C11 SOURCE: TYPE             NUMBER/POINT         POINT INTERVAL
C12     PATTERN:                                LENGTH        WIDTH
C13 SWEEP: START     HZ  END     HZ  LENGTH      MS  CHANNEL NO     TYPE
C14 TAPER: START LENGTH       MS  END LENGTH       MS  TYPE
C15 SPREAD: OFFSET         MAX DISTANCE         GROUP INTERVAL
C16 GEOPHONES: PER GROUP     SPACING     FREQUENCY     MFG          MODEL
C17     PATTERN:                              LENGTH        WIDTH
C18 TRACES SORTED BY: RECORD     CDP     OTHER
C19 AMPLITUDE RECOVERY: NONE      SPHERICAL DIV       AGC    OTHER
```

The SEG-Y EBCDIC Reel Header

```
C20 MAP PROJECTION                      ZONE ID       COORDINATE UNITS
C21 PROCESSING:
C22 PROCESSING:
C23
C24
C25
C26
C27
C28
C29
C30
C31
C32
C33
C34
C35
C36
C37
C38
C39
C40 END EBCDIC
```

The SEG-Y Binary Reel Header

The binary reel header contains much information about the data. Much of this information is optional, that is, the entire header is not required to be valid. In fact, none of it is required to be valid, although some fields are strongly recommended.

```
001 - 004     Job identification number.
005 - 008  *  Line number.
009 - 012  *  Reel number.
013 - 014  *  Number of data traces per record.
015 - 016  *  Number of auxiliary traces per record.
017 - 018  *  Sample interval of this reel's data in microseconds.
019 - 020     Sample interval of original field recording in microseconds.
021 - 022  *  Number of samples per trace for this reel's data.
023 - 024     Number of samples per trace in original field recording.
025 - 026  *  Data sample format code:
                  1 = 32-bit IBM floating point
                  2 = 32-bit fixed-point (integer)
                  3 = 16-bit fixed-point (integer)
                  4 = 32-bit fixed-point with gain code (integer)
027 - 028  *  CDP fold (expected number of data traces per ensemble).
029 - 030     Trace sorting code:
                  1 = as recorded
                  2 = CDP ensemble
                  3 = single fold continuous profile
                  4 = horizontally stacked
```

The SEG-Y Binary Reel Header

```
031 - 032     Vertical sum code (1 = no sum, 2 = two sum, ...)
033 - 034     Sweep frequency at start in Hertz.
035 - 036     Sweep frequency at end in Hertz.
037 - 038     Sweep length in milliseconds.
039 - 040     Sweep type code:
                  1 = linear
                  2 = parabolic
                  3 = exponential
                  4 = other
041 - 042     Trace number of sweep channel.
043 - 044     Sweep trace taper length at start in milliseconds.
045 - 046     Sweep trace taper length at end in milliseconds.
047 - 048     Taper type code:
                  1 = linear
                  2 = cosine squared
                  3 = other
049 - 050     Correlated data traces (1 = no, 2 = yes).
051 - 052     Binary gain recovered (1 = yes, 2 = no).
053 - 054     Amplitude recovery method code:
                  1 = one
                  2 = spherical divergence
                  3 = AGC
                  4 = other
055 - 056  *  Measurement system (1 = meters, 2 = feet).
057 - 058     Impulse signal polarity (increase in pressure or upward
              geophone case movement gives 1=negative or 2=positive number).
```

The SEG-Y Binary Reel Header

```
059 - 060     Vibratory polarity code (seismic lags pilot signal by):
                  1 = 337.5 to 22.5 degrees
                  2 = 22.5 to 67.5 degrees
                  3 = 67.5 to 112.5 degrees
                  4 = 112.5 to 157.5 degrees
                  5 = 157.5 to 202.5 degrees
                  6 = 202.5 to 247.5 degrees
                  7 = 247.5 to 292.5 degrees
                  8 = 292.5 to 337.5 degrees
061 - 400     Unassigned (for optional information).
```

The SEG-Y Trace Header and its keys as seen by Madagascar

The 240-byte binary trace header consists of 2-byte and 4-byte integers in the following layout:

```
1 tracl: trace sequence number within line 0

2 tracr: trace sequence number within reel 4

3 fldr: field record number 8

4 tracf: trace number within field record 12

5 ep: energy source point number 16

6 cdp: CDP ensemble number 20

7 cdpt: trace number within CDP ensemble 24

8 trid: trace identification code:
1 = seismic data    2 = dead            3 = dummy      4 = time break
5 = uphole          6 = sweep           7 = timing     8 = water break
9--, N = optional use (N = 32,767) 28

9 nvs: number of vertically summed traces 30
```

SEG-Y trace header keys

10 nhs: number of horizontally summed traces 32

11 duse: data use:
1 = production 2 = test 34

12 offset: distance from source point to receiver
group (negative if opposite
to direction in which the line was shot) 36

13 gelev: receiver group elevation from sea level (above sea level
is positive) 40

14 selev: source elevation from sea level (above sea level is
positive) 44

15 sdepth: source depth (positive) 48

16 gdel: datum elevation at receiver group 52

17 sdel: datum elevation at source 56

18 swdep: water depth at source 60

19 gwdep: water depth at receiver group 64

20 scalel: scale factor for previous 7 entries with value plus or
minus 10 to the
power 0, 1, 2, 3, or 4 (if positive, multiply, if negative divide) 68

SEG-Y trace header keys

21 scalco: scale factor for next 4 entries with value plus or
minus 10 to the
power 0, 1, 2, 3, or 4 (if positive, multiply, if negative divide) 70

22 sx: X source coordinate 72

23 sy: Y source coordinate 76

24 gx: X group coordinate 80

25 gy: Y group coordinate 84

26 counit: coordinate units code for previous four entries
1 = length (meters or feet)
2 = seconds of arc (in this case, the X values are unsigned longitude and the Y
values
are latitude, a positive value designates the number of seconds east of
Greenwich
or north of the equator 88

27 wevel: weathering velocity 90

28 swevel: subweathering velocity 92

29 sut: uphole time at source 94

30 gut: uphole time at receiver group 96

SEG-Y trace header keys

31 sstat: source static correction 98

32 gstat: group static correction 100

33 tstat: total static applied 102

34 laga: lag time A, time in ms between end of 240-byte trace
identification header and time
break, positive if time break occurs after end of header, time break is defined
as
the initiation pulse which maybe recorded on an auxiliary trace or as otherwise
specified by the recording system 104

35 lagb: lag time B, time in ms between the time break and the
initiation time of the energy
source, may be positive or negative 106

36 delrt: delay recording time, time in ms between initiation
time
of energy source and time
when recording of data samples begins (for deep water work if recording does not
start at zero time) 108

37 muts: mute time-start 110

38 mute: mute time-end 112

SEG-Y trace header keys

39 ns: number of samples in this trace 114

40 dt: sample interval, in micro-seconds 116

41 gain: gain type of field instruments code:
1 = fixed 2 = binary 3 = floating point 4 -- N = optional use
118

42 igc: instrument gain constant 120

43 igi: instrument early or initial gain 122

44 corr: correlated:
1 = no 2 = yes 124

45 sfs: sweep frequency at start 126

46 sfe: sweep frequency at end 128

47 slen: sweep length in ms 130

48 styp: sweep type code:
1 = linear 2 = cos-squared 3 = other 132

49 stas: sweep trace length at start in ms 134

50 stae: sweep trace length at end in ms 136

SEG-Y trace header keys

51 tatyp: taper type: 1=linear, 2=cos^2, 3=other 138

52 afilf: alias filter frequency if used 140

53 afils: alias filter slope 142

54 nofilf: notch filter frequency if used 144

55 nofils: notch filter slope 146

56 lcf: low cut frequency if used 148

57 hcf: high cut frequncy if used 150

58 lcs: low cut slope 152

59 hcs: high cut slope 154

60 year: year data recorded 156

61 day: day of year 158

62 hour: hour of day (24 hour clock) 160

63 minute: minute of hour 162

64 sec: second of minute 164

SEG-Y trace header keys

65 timbas: time basis code:
1 = local 2 = GMT 3 = other 166

66 trwf: trace weighting factor, defined as 1/2^N volts for the least sigificant bit 168

67 grnors: geophone group number of roll switch position one 170

68 grnofr: geophone group number of trace one within original field record 172

69 grnlof: geophone group number of last trace within original field record 174

70 gaps: gap size (total number of groups dropped) 176

71 otrav: overtravel taper code:
1 = down (or behind) 2 = up (or ahead) 178

72 cdpx: X coordinate of CDP 180

73 cdpy: Y coordinate of CDP 184

74 iline: in-line number 188

75 xline: cross-line number 192

76 shnum: shotpoint number 196

SEG-Y trace header keys

77 shsca: shotpoint scalar 200

78 tval: trace value meas. 202

79 tconst4: transduction const 204

80 tconst2: transduction const 208

81 tunits: transduction units 210

82 device: device identifier 212

83 tscalar: time scalar 214

84 stype: source type 216

85 sendir: source energy dir. 218

86 unknown: unknown 222

87 smeas4: source measurement 224

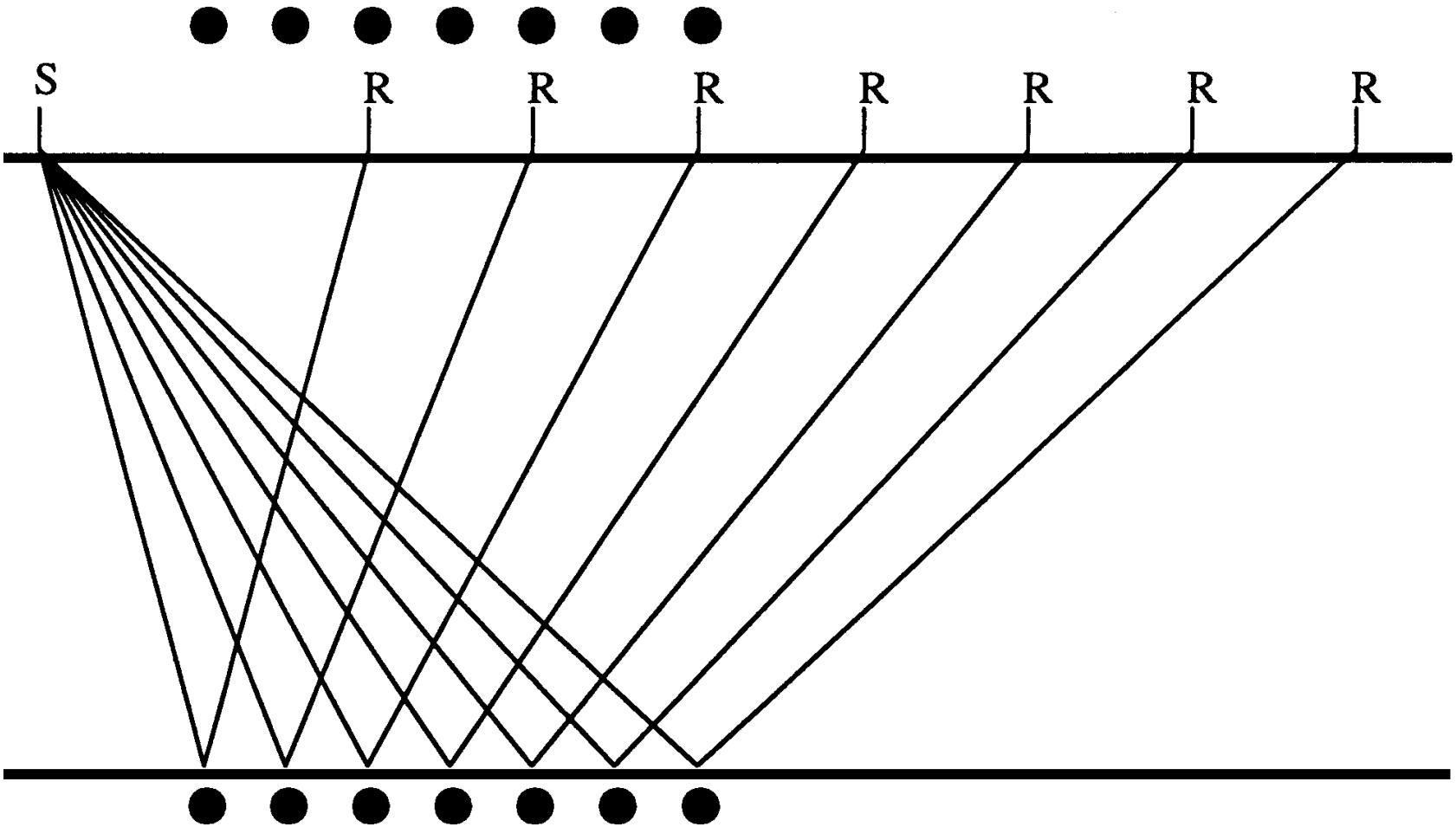
88 smeas2: source measurement 228

89 smeasu: source measurement unit 230

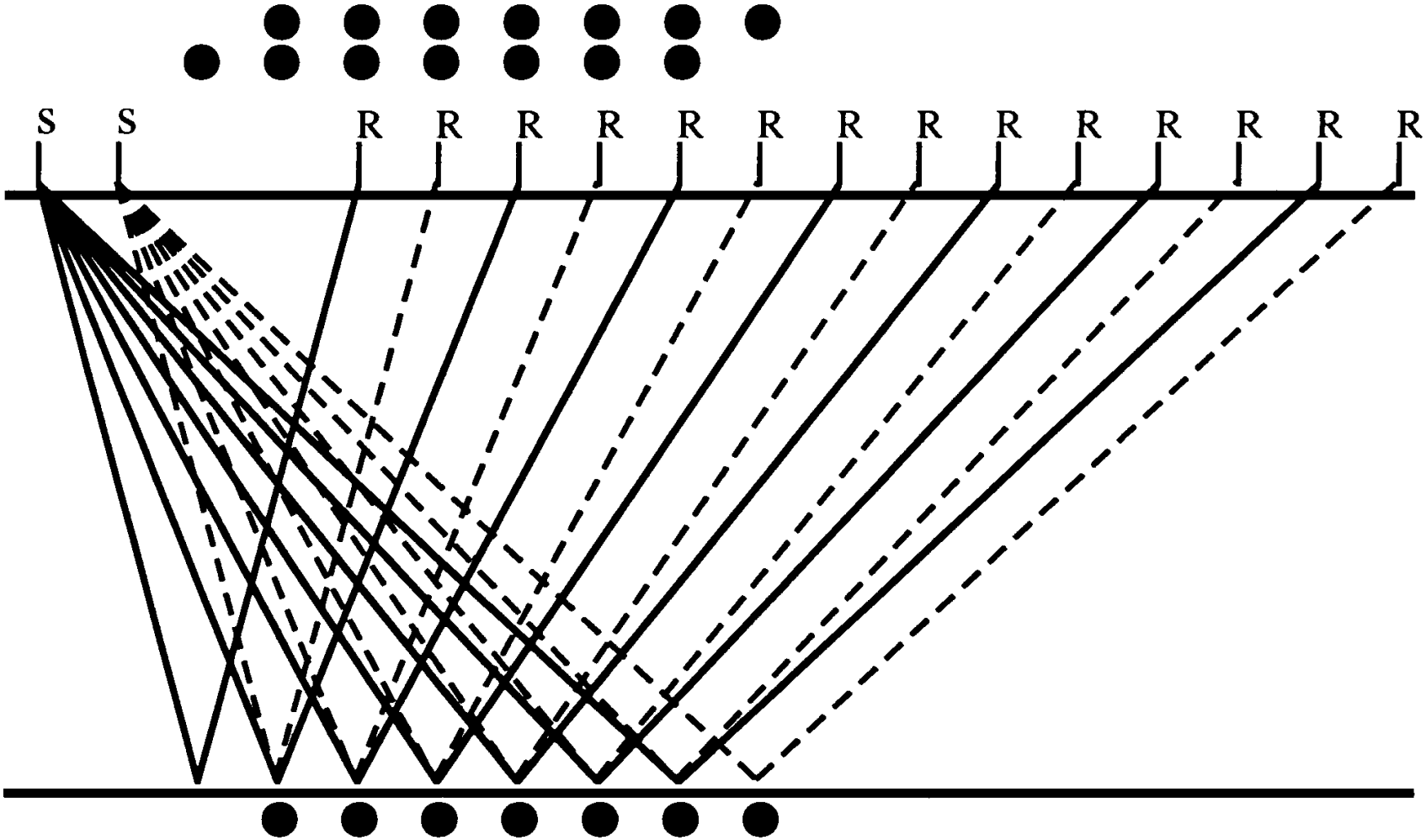
90 unass1: unassigned 232

91 unass2: unassigned 236

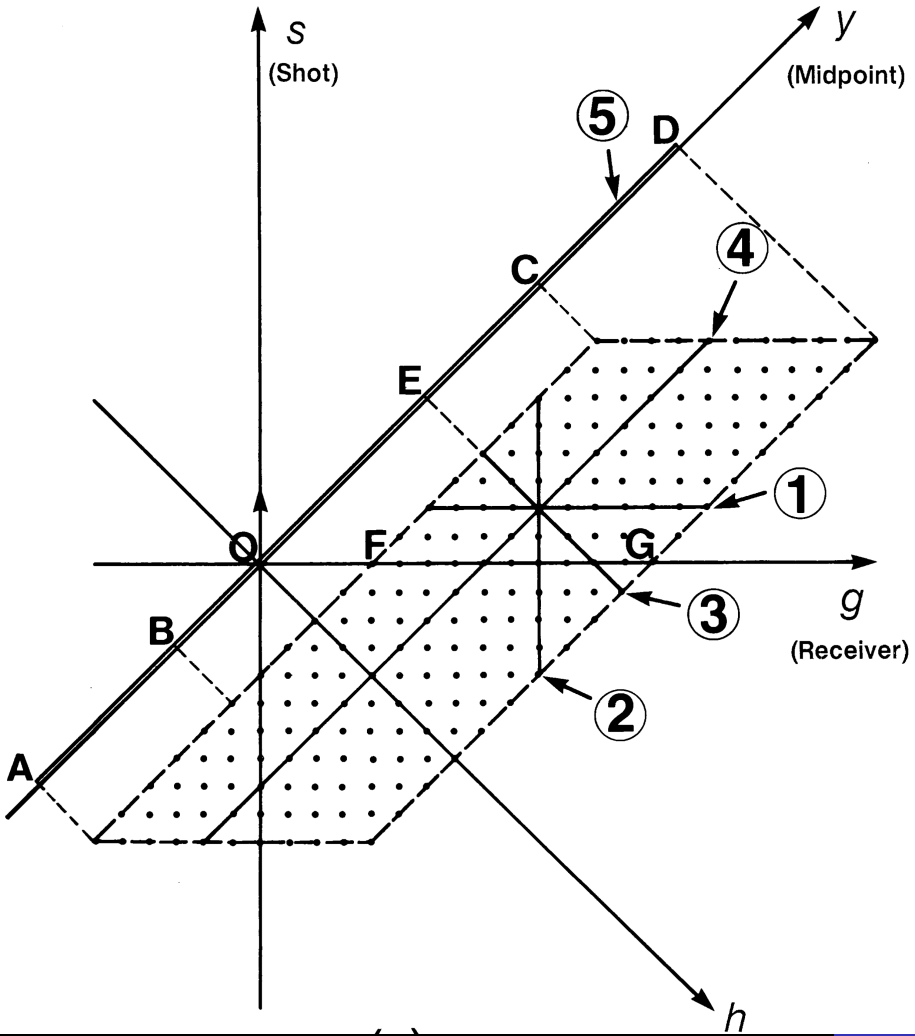
Aquisition

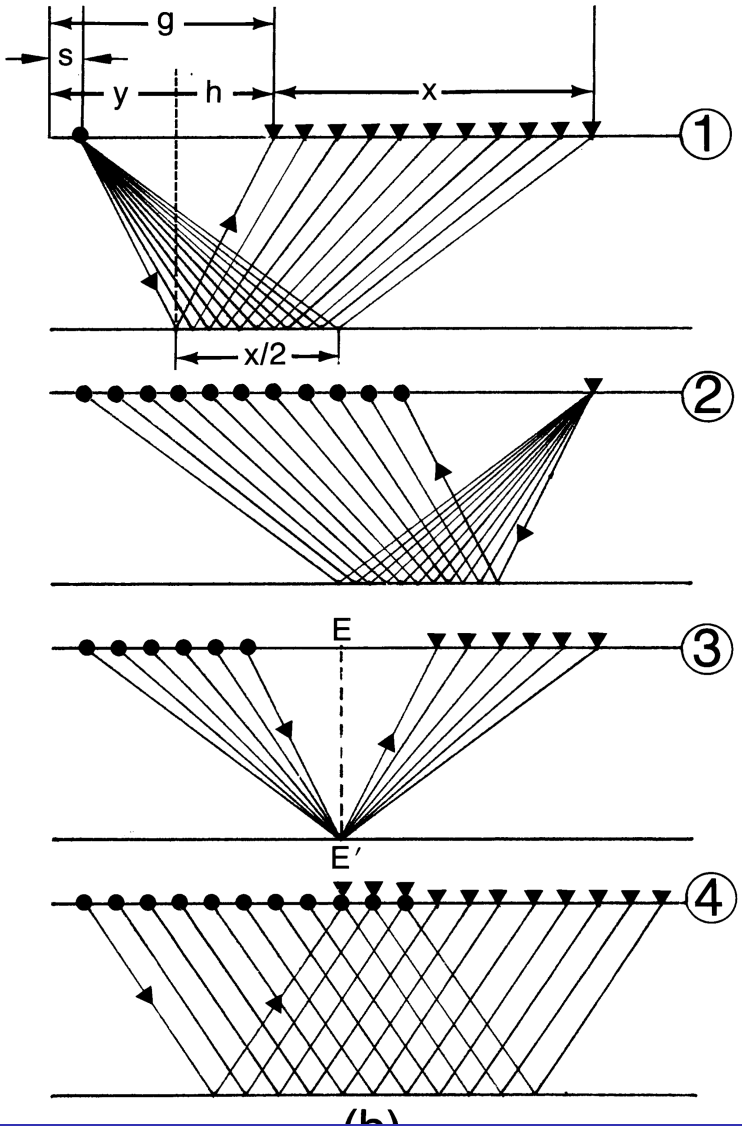


Aquisition

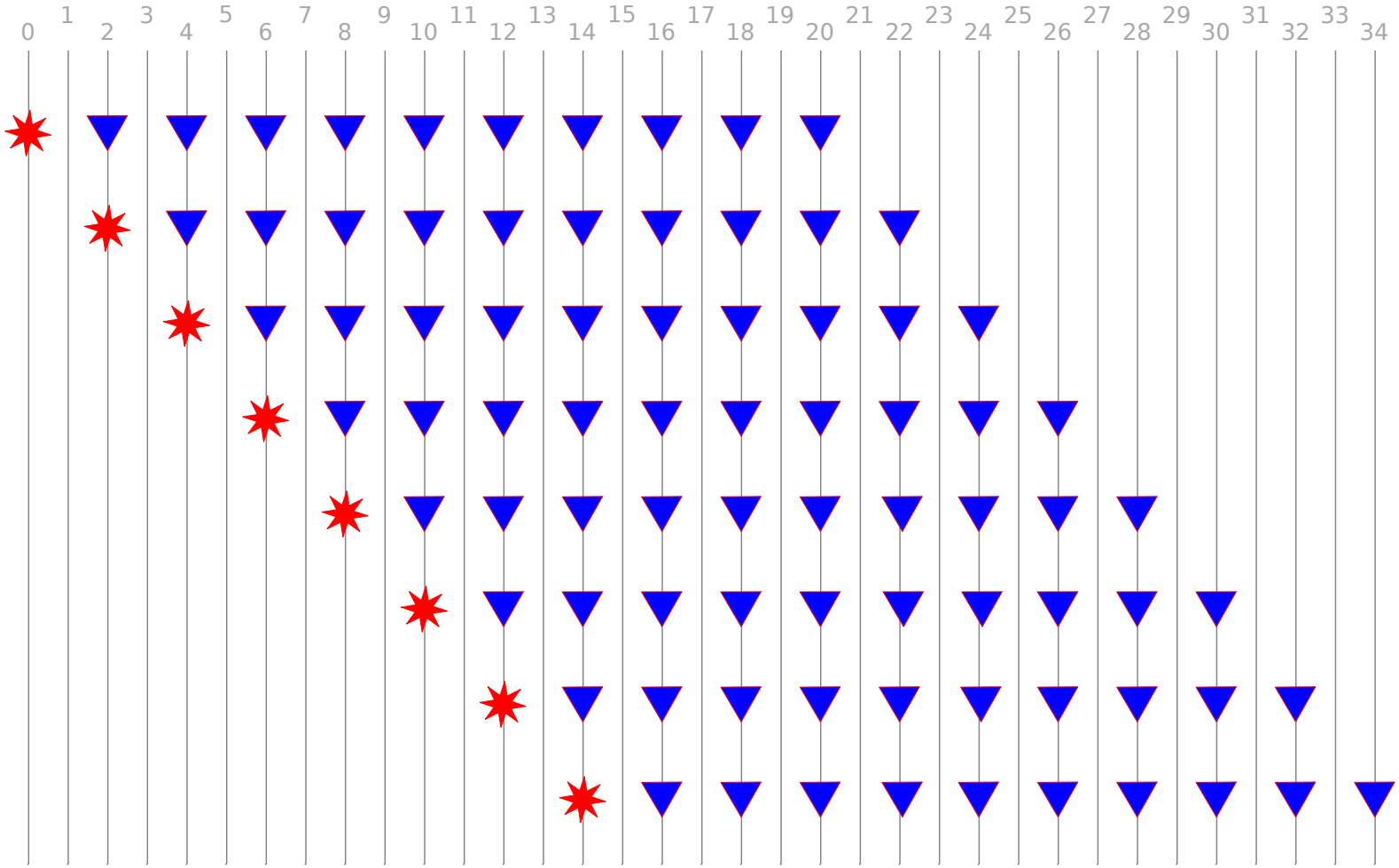


Stacking chart/diagram

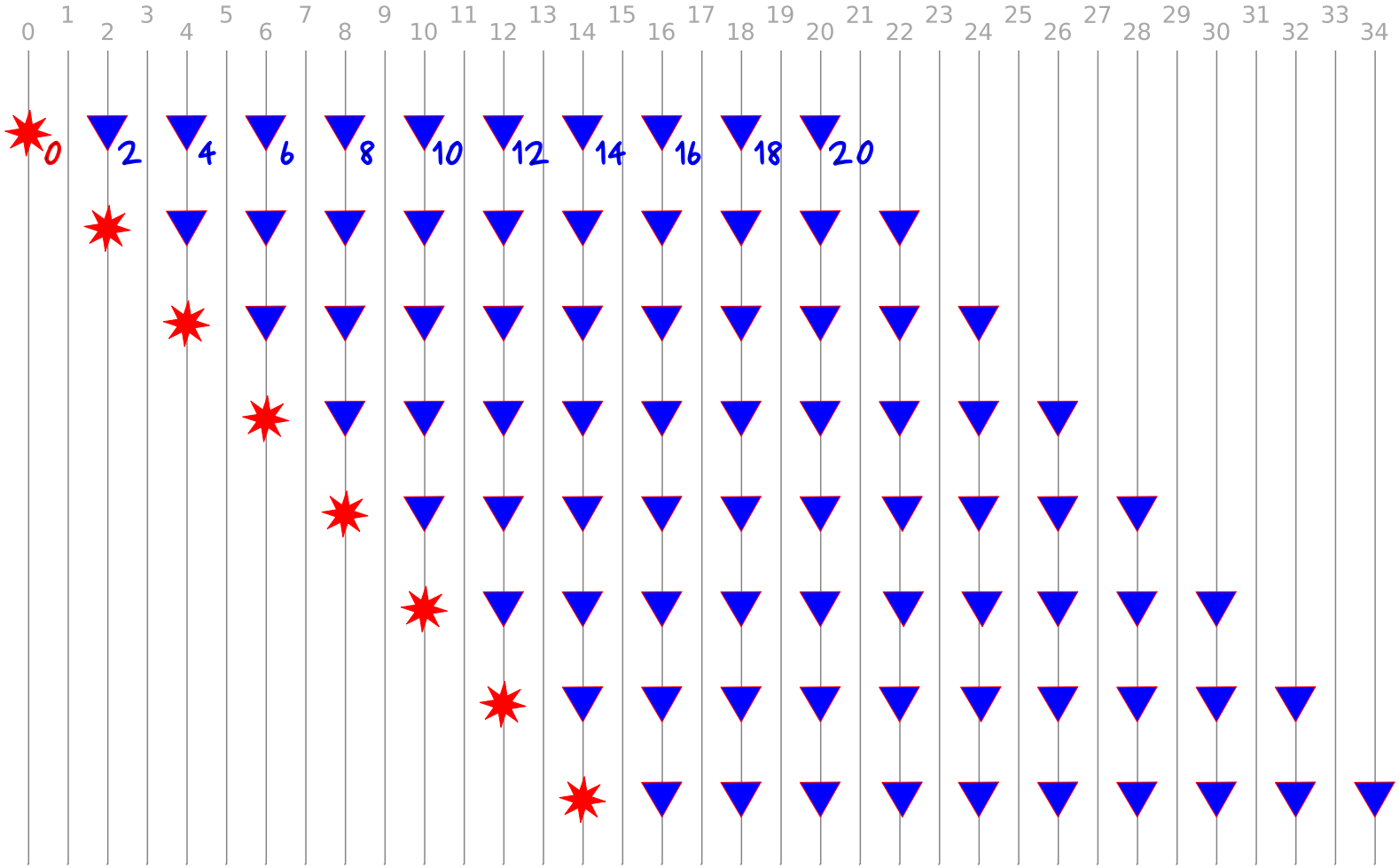




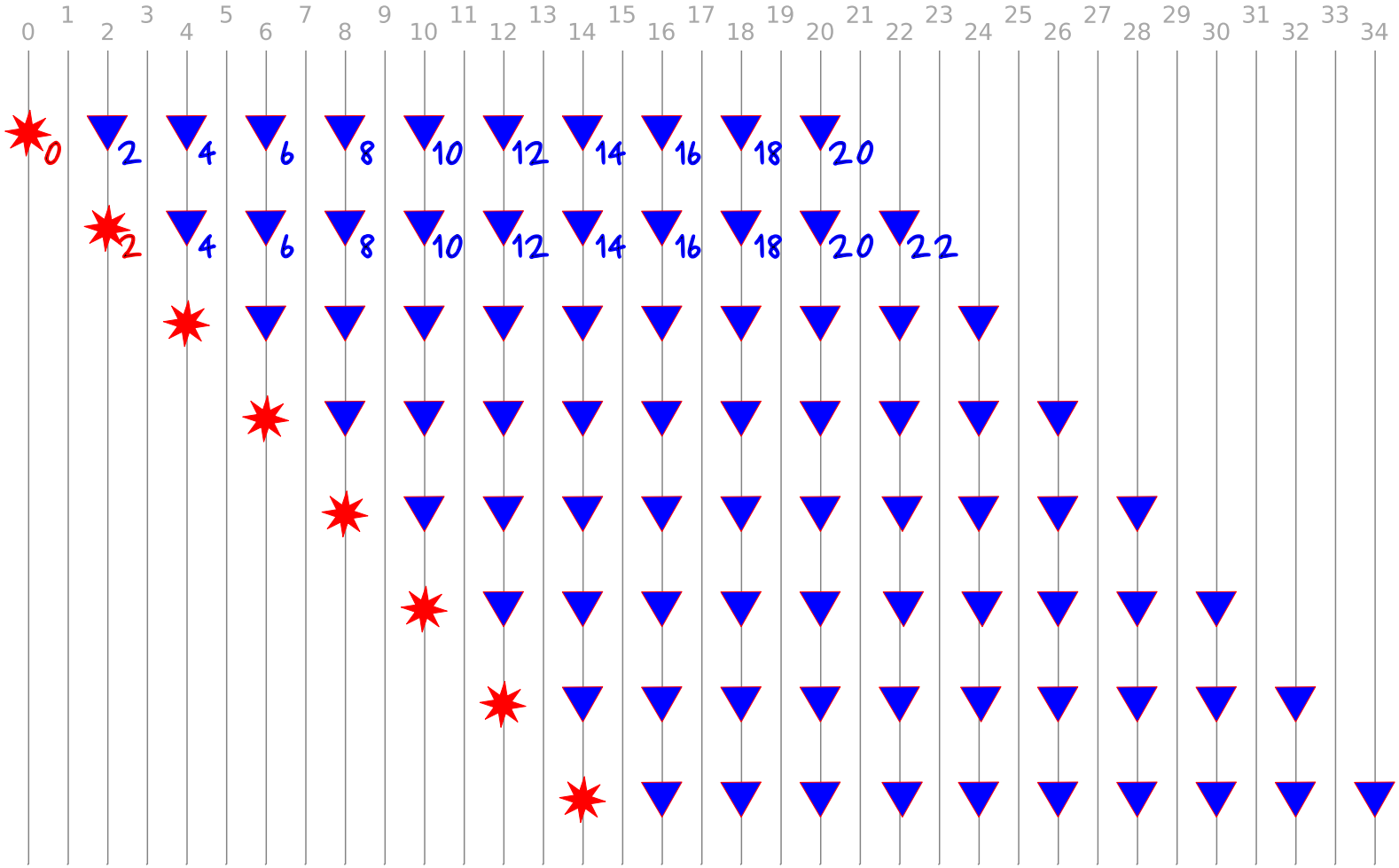
Gather sorting



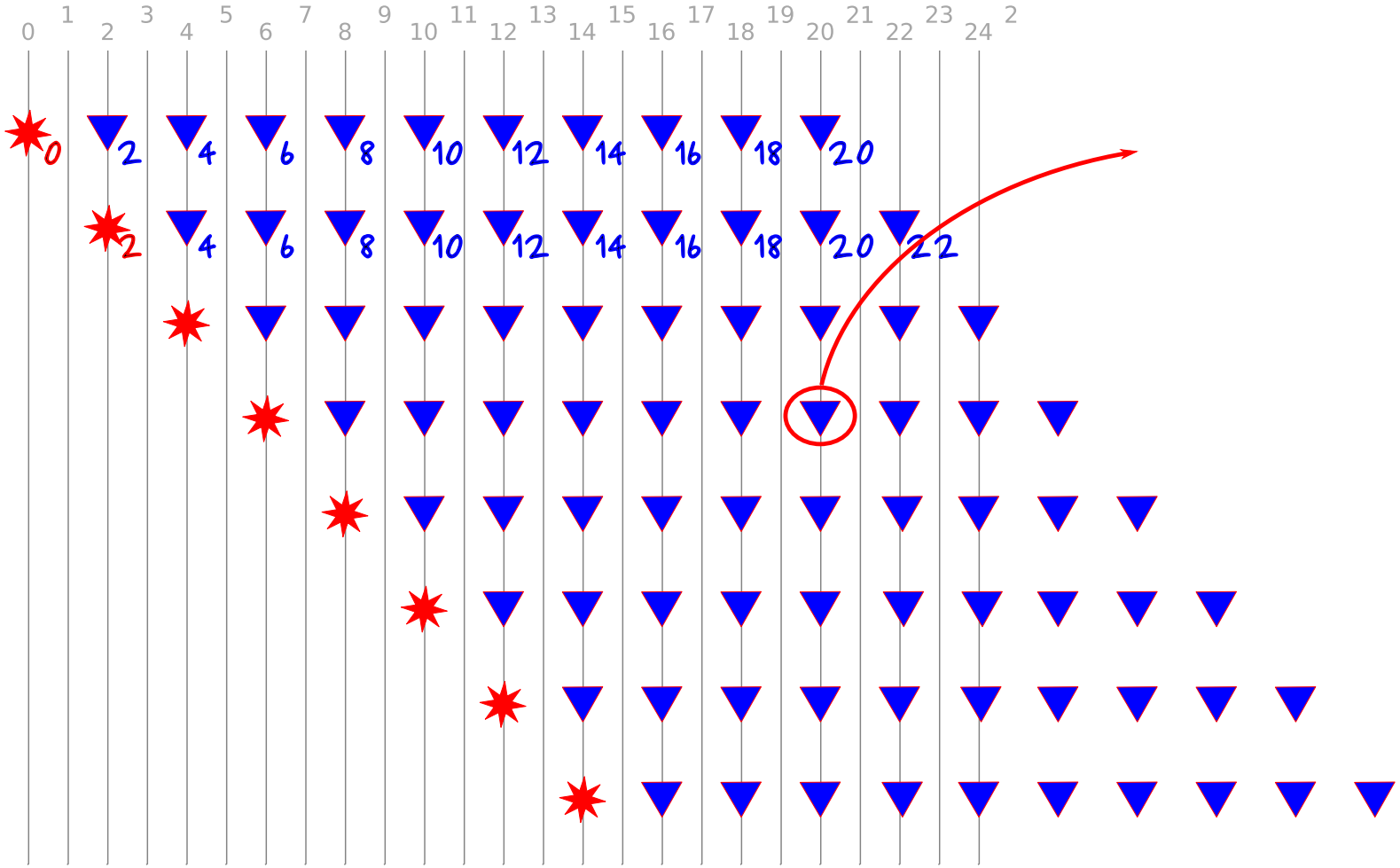
Gather sorting



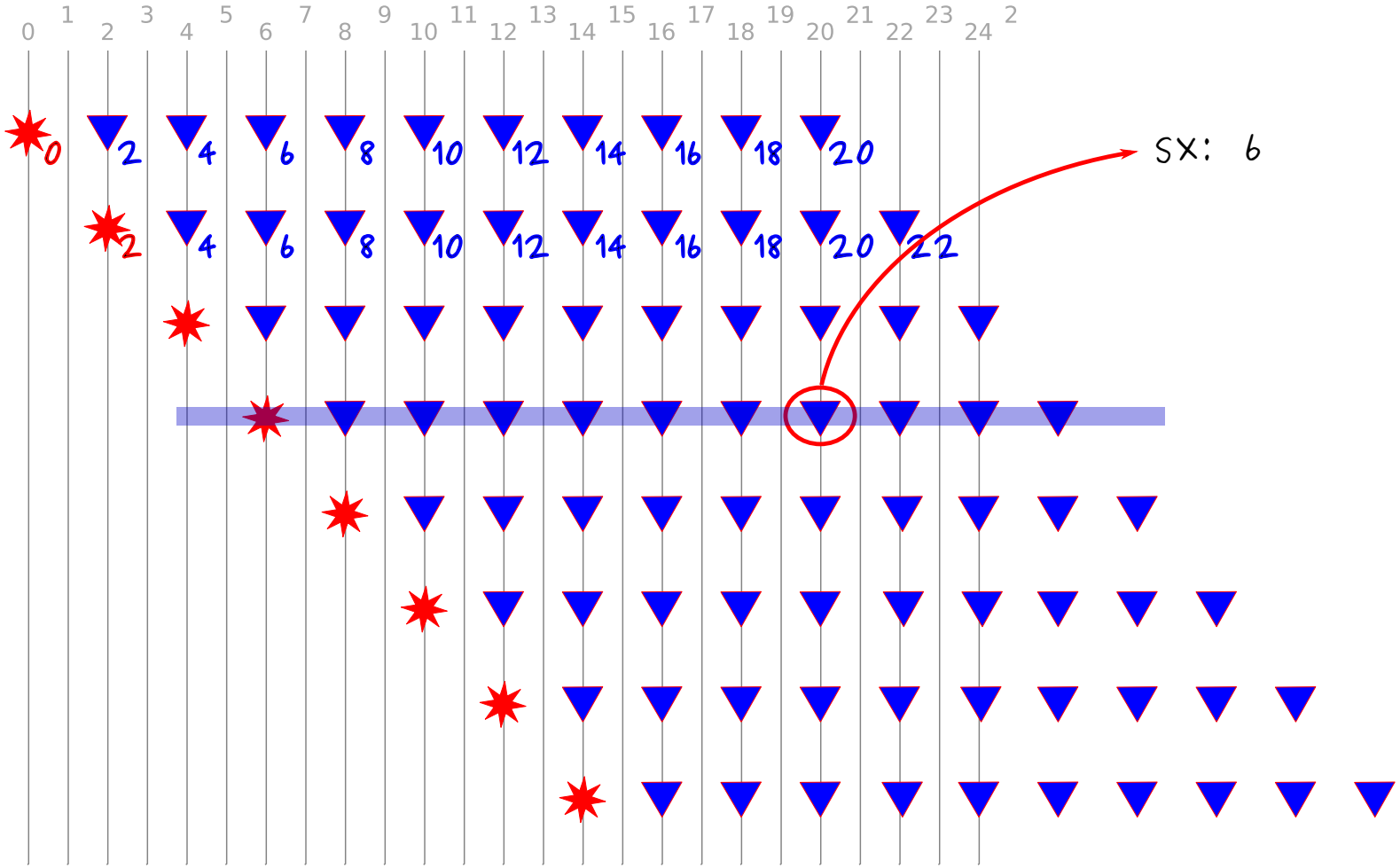
Gather sorting



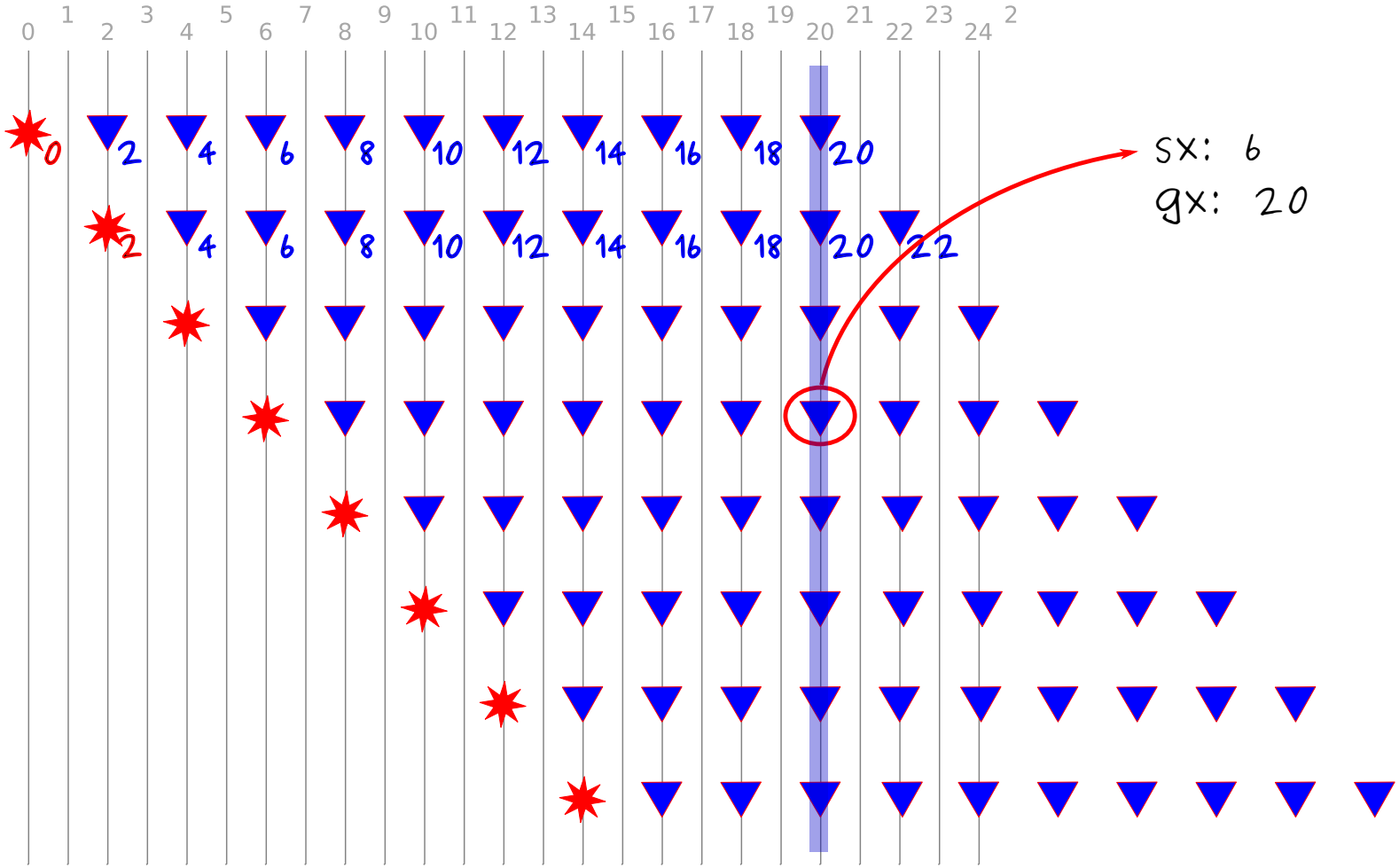
Gather sorting



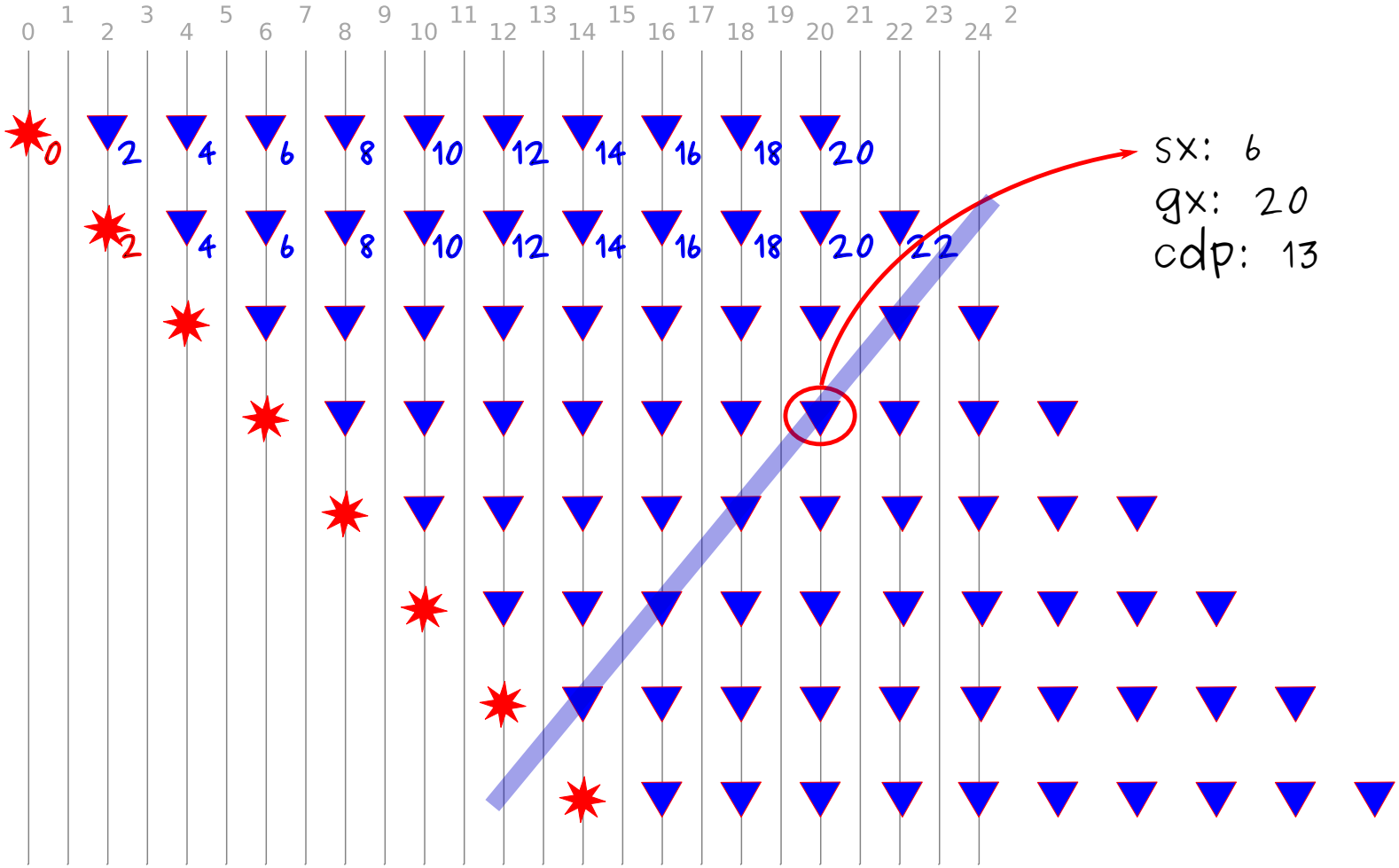
Gather sorting



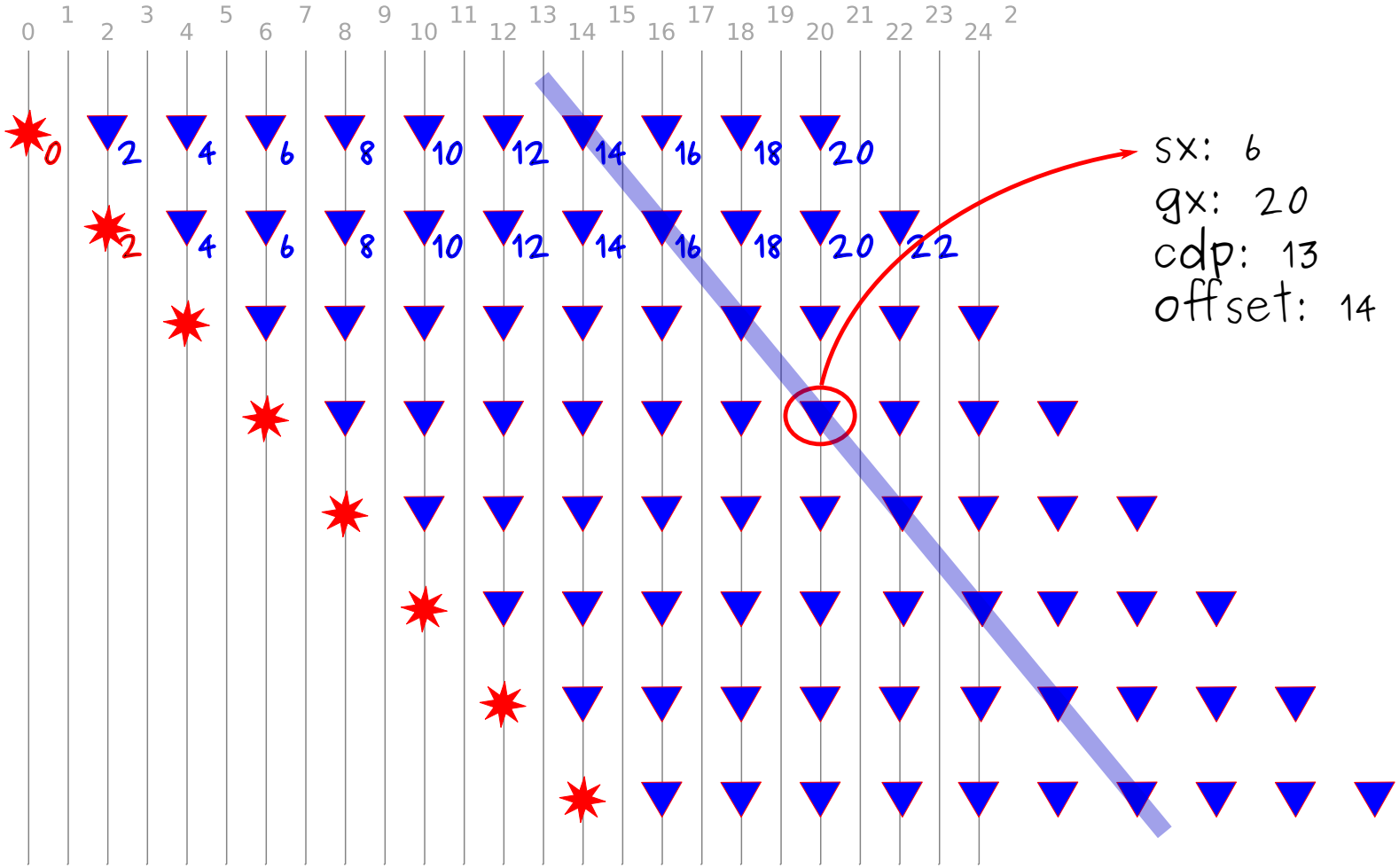
Gather sorting



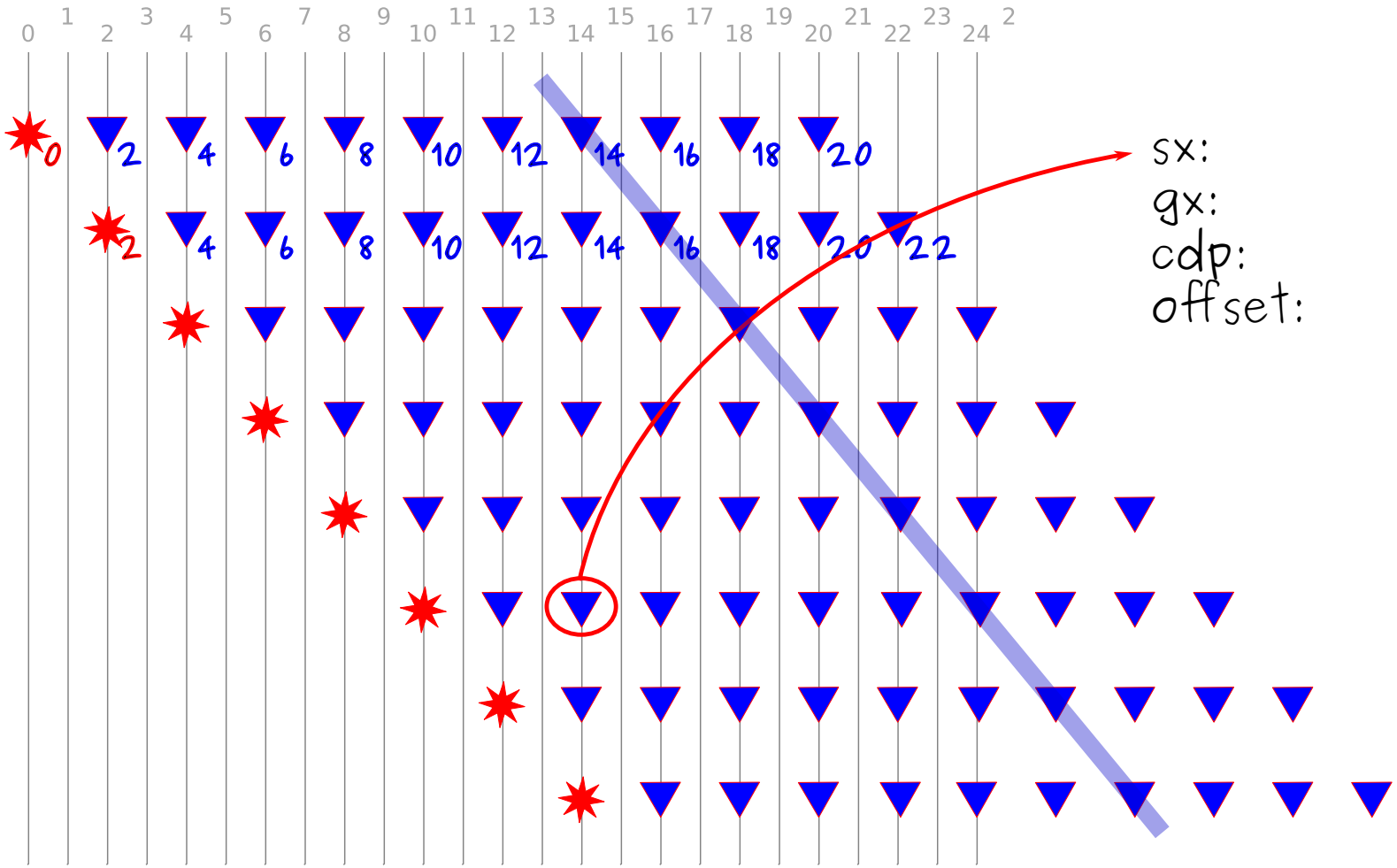
Gather sorting



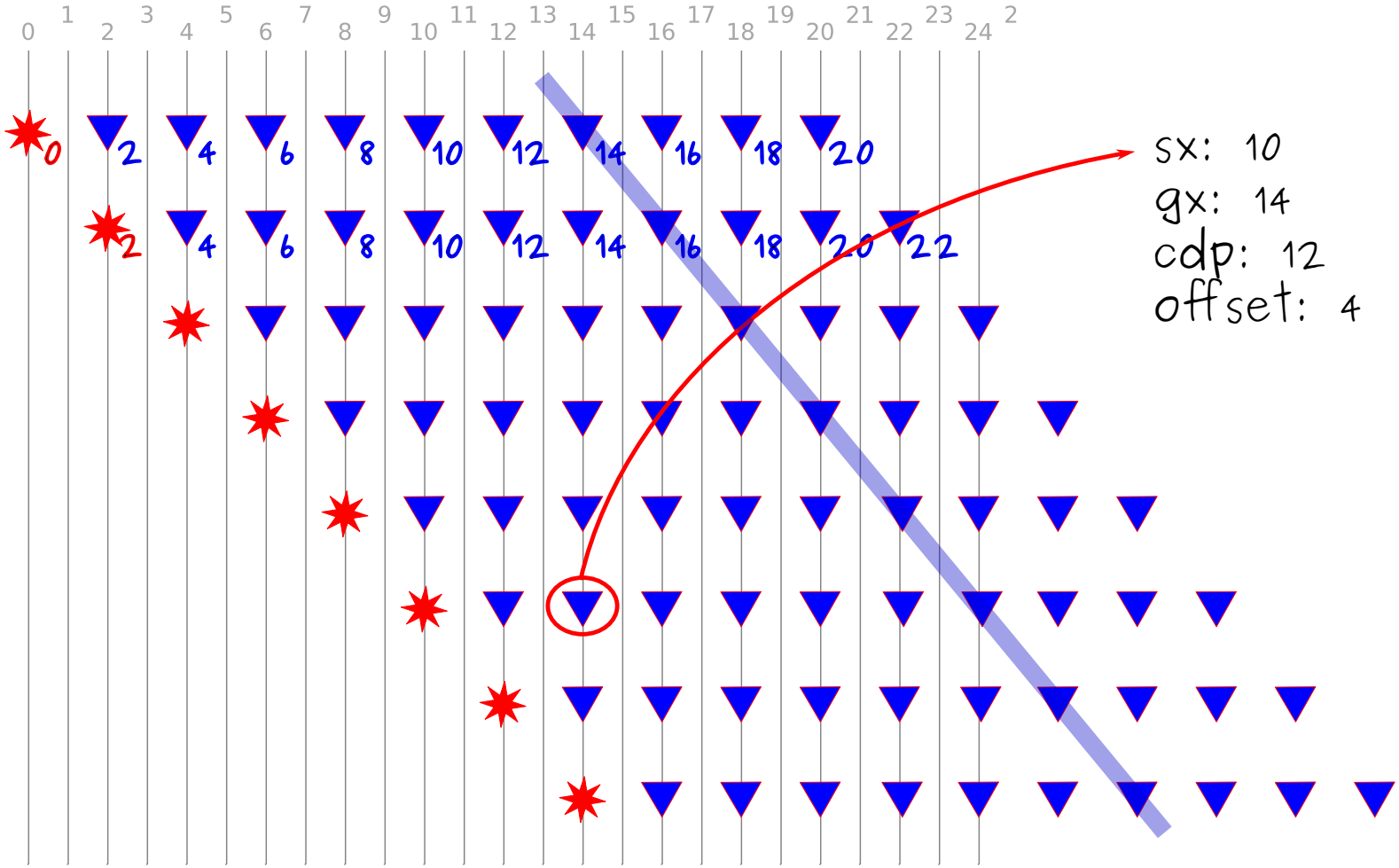
Gather sorting



Gather sorting



Gather sorting



SEG-Y to RSF

Reading SEG-Y

```
sfsegyread mask=msk.rsf > out.rsf tfile=hdr.rsf verb=n su=
suxdr=n endian=y n2=0 format=segyformat (bhead) ns=segyns (bhead)
tape= read= hfile= bfile=
```

```
bash$ sfsegyread
```

```
NAME
        sfsegyread
DESCRIPTION
        Convert a SEG-Y or SU dataset to RSF.
SYNOPSIS
        sfsegyread mask=msk.rsf > out.rsf tfile=hdr.rsf verb=n su= suxdr=n endian=y n2=0 for
mat=segyformat (bhead) ns=segyns (bhead) tape= read= hfile= bfile=
COMMENTS

        Data headers and trace headers are separated from the data.

        "suread" is equivalent to "segyread su=y"

        SEGY key names:

        tracl: trace sequence number within line 0
...
```

SEG-Y to RSF

Reading SEG-Y

```
sfsegyread mask=msk.rsf > out.rsf tfile=hdr.rsf verb=n su=
suxdr=n endian=y n2=0 format=segyformat (bhead) ns=segyns (bhead)
tape= read= hfile= bfile=
```

```
PARAMETERS
        string  bfile=  output binary data header file
        bool    endian=y [y/n]  Whether to automatically estimate endianness or not
        int     format=segyformat (bhead) [1,2,3,5]     Data format. The default is taken from binary
header.
                   1 is IBM floating point
                   2 is 4-byte integer
                   3 is 2-byte integer
                   5 is IEEE floating point

        string  hfile=  output text data header file
        string  mask=   optional header mask for reading only selected traces (auxiliary input file name)
        int     n2=0    number of traces to read (if 0, read all traces)
        int     ns=segyns (bhead)       Number of samples. The default is taken from binary header
        string  read=   what to read: h - header, d - data, b - both (default)
        bool    su= [y/n]       y if input is SU, n if input is SEGY
        bool    suxdr=n [y/n]   y, SU has XDR support
        string  tape=   input data
        string  tfile=  output trace header file (auxiliary output file name)
        bool    verb=n [y/n]    Verbosity flag
```

RSF to SEG-Y

Writing SEG-Y

```
sfsegywrite < in.rsf tfile=hdr.rsf verb=n endian=y su= suxdr=n
tape= hfile= bfile=
```

```
bash$ sfsegywrite
```

```
NAME
        sfsegywrite
DESCRIPTION
        Convert an RSF dataset to SEGY or SU.
SYNOPSIS
        sfsegywrite < in.rsf tfile=hdr.rsf verb=n endian=y su= suxdr=n suxdr=n
tape= hfile= bfile=
COMMENTS

        Merges trace headers with data.

        "suwrite" is equivalent to "segywrite su=y"

PARAMETERS
        string  bfile=  input binary data header file
        bool    endian=y [y/n]  Whether to automatically estimate endianness or
not
        string  hfile=  input text data header file
        bool    su= [y/n]       y if input is SU, n if input is SEGY
        bool    suxdr=n [y/n]   y, SU has XDR support
```

Conversion with ASCII

ASCII to RSF

```
echo in=in.asc data_format=ascii_float | sfdd form=native >
out.rsf
```

```
bash$ cat g.asc
```

```
1 2 3 4 5 6
```

```
bash$ echo in=g.asc n1=3 o1=1 d1=1 n2=2 o2=1 d2=1
data_format=ascii_int | sfdd form=native type=float > g.rsf
bash$ sfin g.rsf
```

```
g.rsf:
    in="/home/dlmacbr/rsfdata/g.rsf@"
    esize=4 type=float form=native
    n1=3            d1=1            o1=1
    n2=2            d2=1            o2=1
        6 elements 24 bytes
```

Conversion with ASCII

RSF to ASCII

```
sfdd form=ascii out=out.asc < in.rsf > /dev/null
```

```
bash$ sfdisfil < Vel.rsf | less
bash$ sfdd < Vel.rsf form=ascii out=Vel.asc > /dev/null
bash$ gedit Vel.asc &
bash$ sfattr < Vel.rsf
```

```
*****************************************
     rms =       2023.69
    mean =       1976.82
  2-norm =        703364
variance =        187509
 std dev =       433.023
     max =       2599.99 at 201 253
     min =       1499.57 at 1 1
nonzero samples = 120801
  total samples = 120801
*****************************************
```

```
bash$ sfdisfil < Vel.rsf format=%13.8g | less
```