Wander in MLD (CTBI) encoded links

Pete Anslow, Nortel Networks

IEEE P802.3ba, Portland, January 2008

Introduction

The “MLD” proposal (aka CTBI) for striping data across multiple lanes in 40 and 100 Gb Ethernet generates a number of 64B/66B encoded virtual lanes and then bit interleaves them to form the required number of electrical or optical lanes. See gustlin_01_0108.pdf.

In order to be able to identify and re-align the lanes in the receiver a periodic unscrambled lane alignment marker is added to each of the virtual lanes.

This contribution analyses the effect of the proposed alignment markers on baseline wander and variation in clock content.

Baseline wander

In order to be able to compare the effect of various encoding proposals on the low frequency and clock content characteristics of the signals, it is useful to define a low frequency content metric (Baseline Wander) and also a clock content metric (Clock Wander).

Since it is highly likely that these signals will be AC coupled at some point, the baseline wander has been analysed by calculating the amount of offset due to AC coupling.

In order to make use of recent analysis done within OIF this is the same metric as was used in the OIF white paper:

http://www.oiforum.com/public/documents/OIF_WP_CEI_Short_Stress_Patterns.pdf

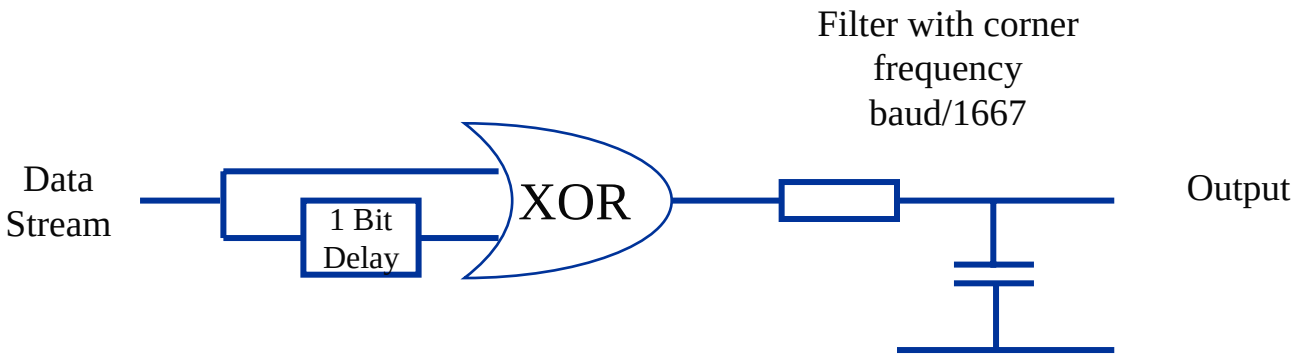
This is defined as:

Baseline Wander is the instantaneous offset (in %) in the signal generated by AC coupling at the bit rate / 10,000.

Clock wander

Since many of the links that 40 and 100 GbE traverse require the clock to be extracted from the lane data at the receiver, a second metric to assess the time variation of the clock content is required. The function used in the OIF white paper (and used in this contribution) is:

Create a function which is a 1 for a transition and a 0 for no transition and then filter the resulting sequence with a corner frequency of baud/1667



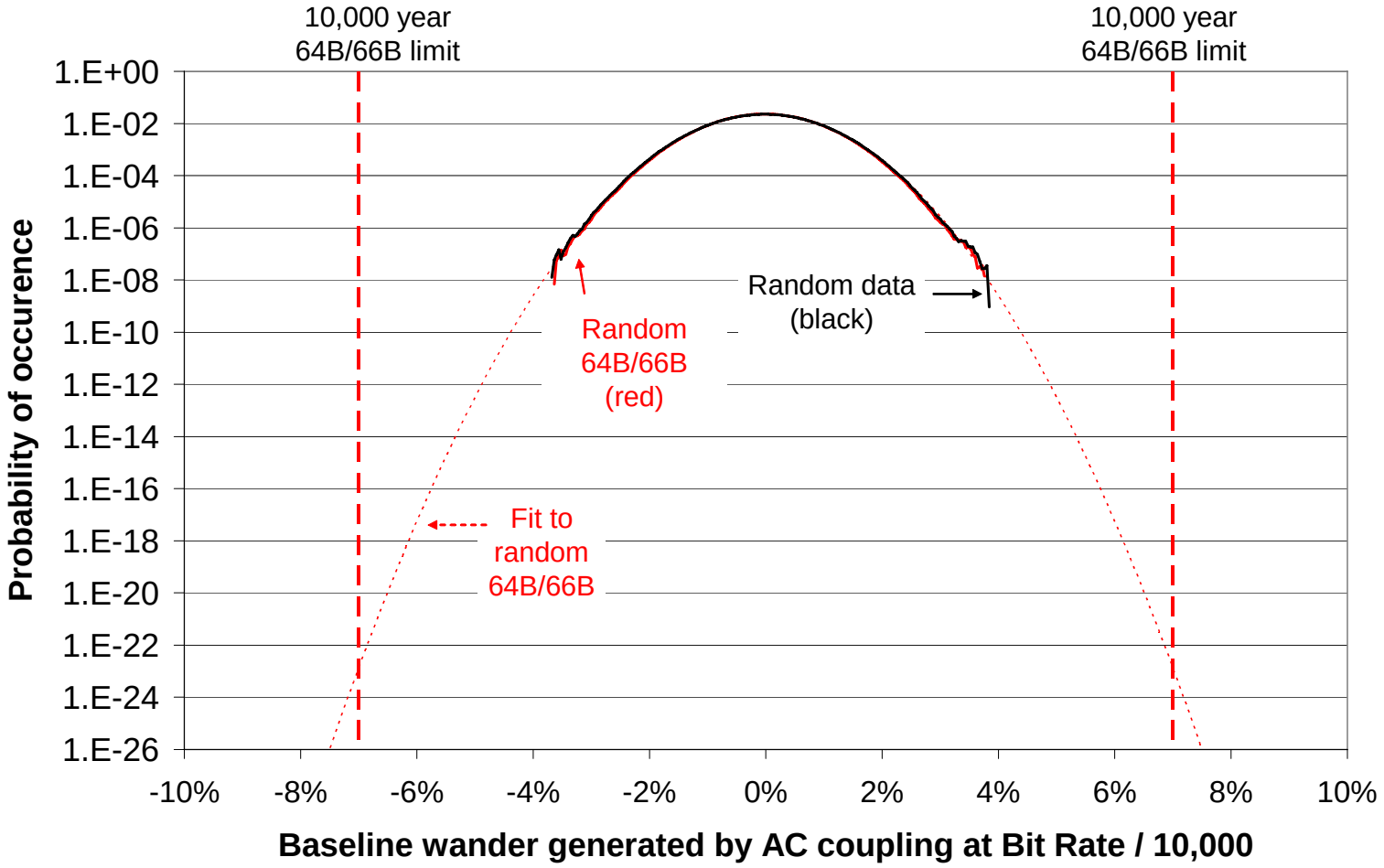
64B/66B coded data

To establish a baseline against which the new proposals can be compared, a simple simulation of 64B/66B coded random data was analysed for Baseline Wander and Clock Wander.

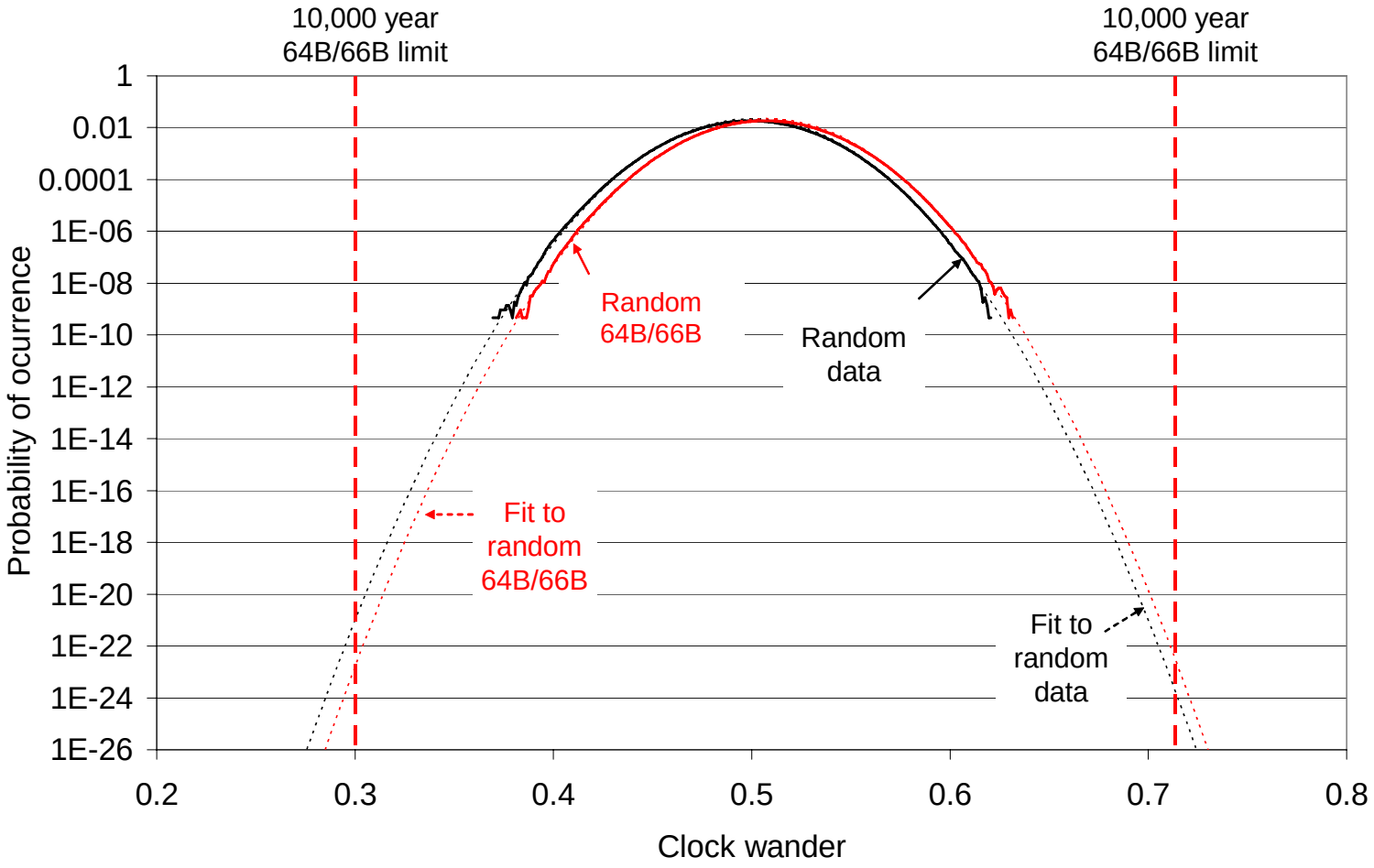
This data is presented in the next two slides as Probability Density Functions (PDFs) i.e. the probability at any instant that the baseline wander or clock wander exceeds a given value. The solid curves are the results from 2×10^9 bits and the dotted curves are fitted to this data.

Also plotted for comparison are the PDFs for simple random binary and the vertical dashed lines give the wander value that will be exceeded only once in 10,000 years for the 64B/66B coded random data at 100 Gbit/s.

Baseline wander PDF of 64B/66B coded random



Clock wander PDF of 64B/66B coded random



Worst case MLD

The worst case for the MLD proposal is for 100 GbE with 20 virtual lanes when all of those lanes are bit interleaved in to a single serial stream as they would be in a possible future 100 GbE serial PHY.

In this case every bit that is the same across **all** of the virtual lanes turns in to 20 identical bits in the serial case. Therefore choose to fill the alignment block as below:

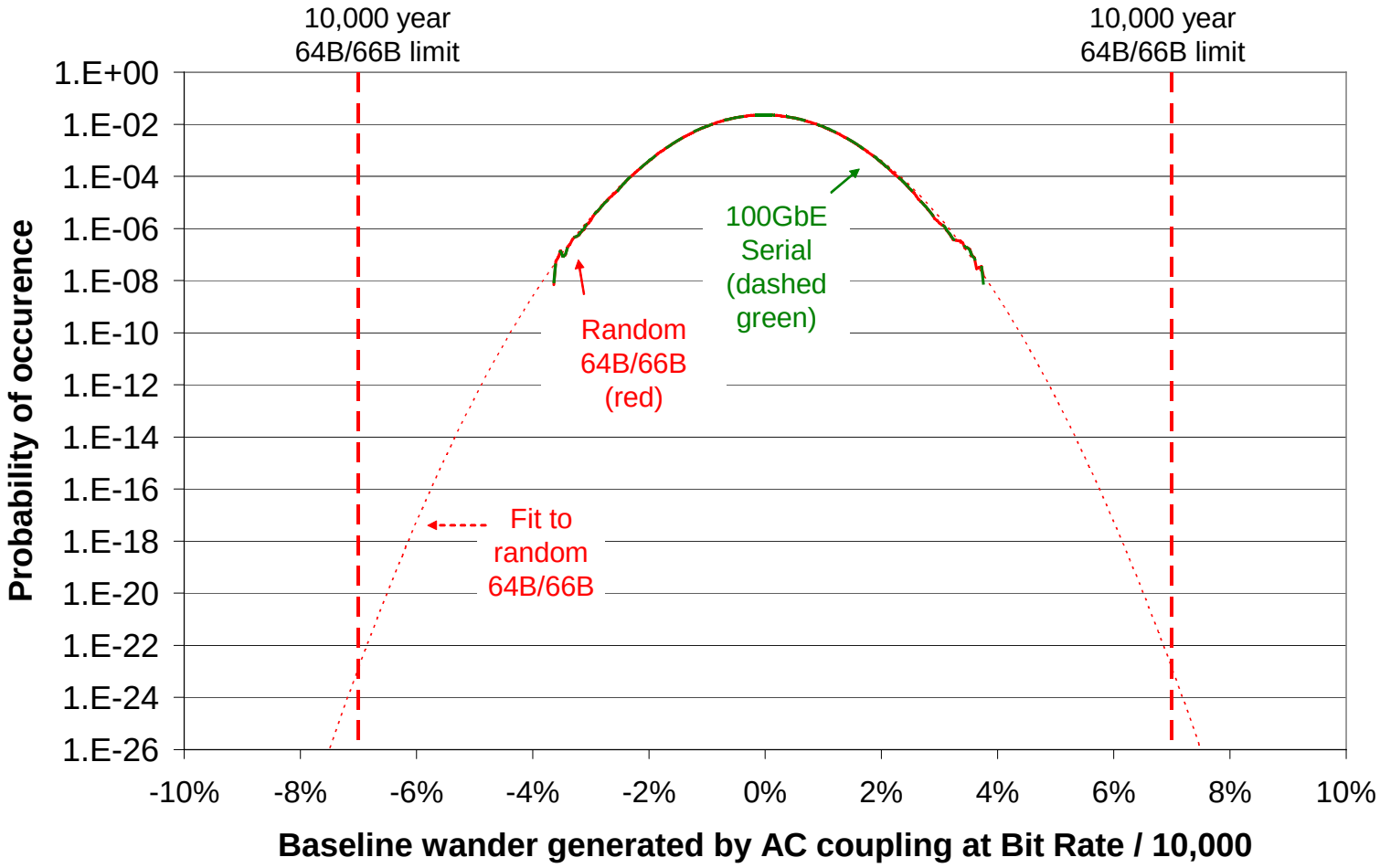
Proposed Alignment Block

10	$\overline{VL}$	VL	$\overline{VL}$	VL	$\overline{VL}$	VL	$\overline{VL}$	VL

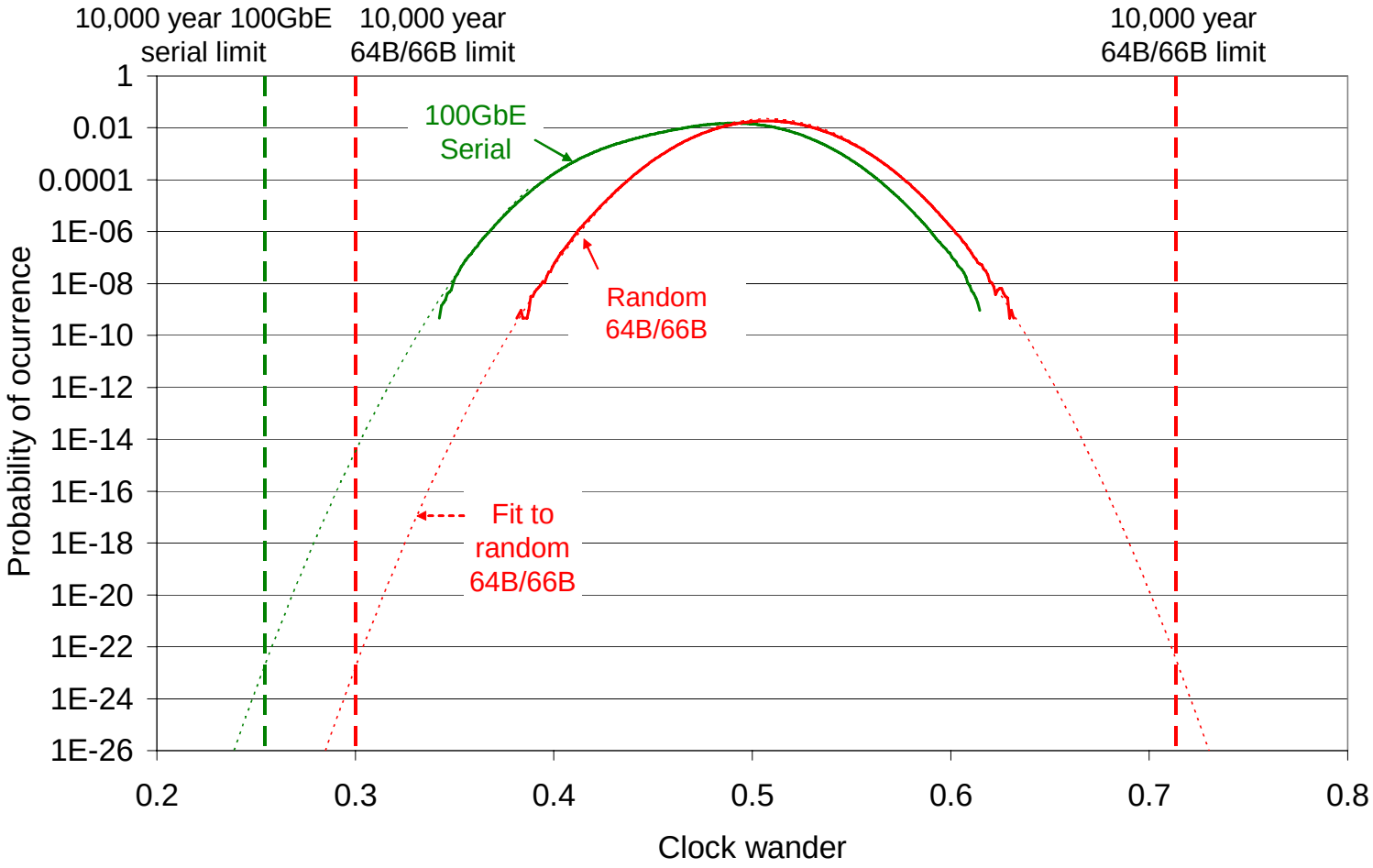
Where VL for lanes 0 to 19 is 54, CE, E9, 63, 7B, 5B, 24, 70, BE, 57, 34, 47, 14, 30, 40, FE, A9, 9D, D2, C6 respectively.

The resulting, PDFs are shown on the next two charts.

Baseline wander PDF for 100G serial MLD



Clock wander PDF for 100G serial MLD



Results

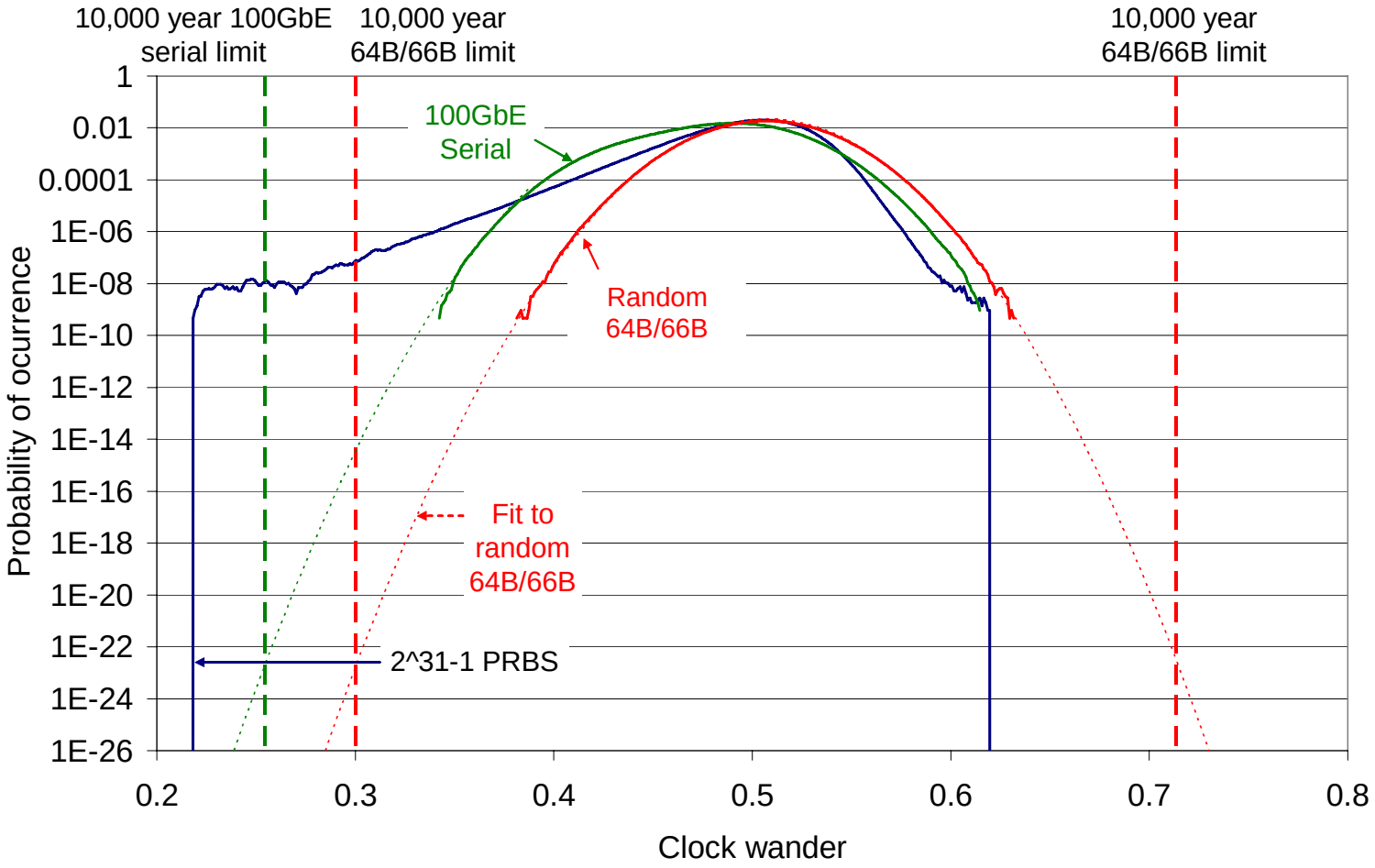
The baseline wander PDF curve for 100GbE serial with the proposed lane markers is almost identical to that of 64B/66B coded random data.

The clock wander PDF characteristic shows a shift towards lower clock content compared to 64B/66B coded random data. This is due to the alignment of all of the unscrambled “01” bits at the start of the data blocks and the “10” bits at the start of the lane alignment blocks.

This simulation is the worst case for this as any skew between the lanes will reduce the degree of alignment.

To put this shift in to perspective, the PDF of a commonly specified test pattern (a pseudo-random binary sequence of length 2^31-1 bits) is also shown on the next slide for comparison. This has a minimum clock wander value of 0.219 compared to the value expected for 10,000 years of 100GbE serial of 0.254.

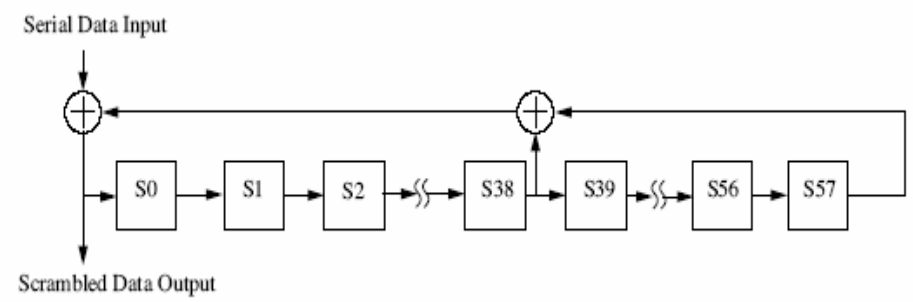
Clock wander PDF for 100G serial MLD



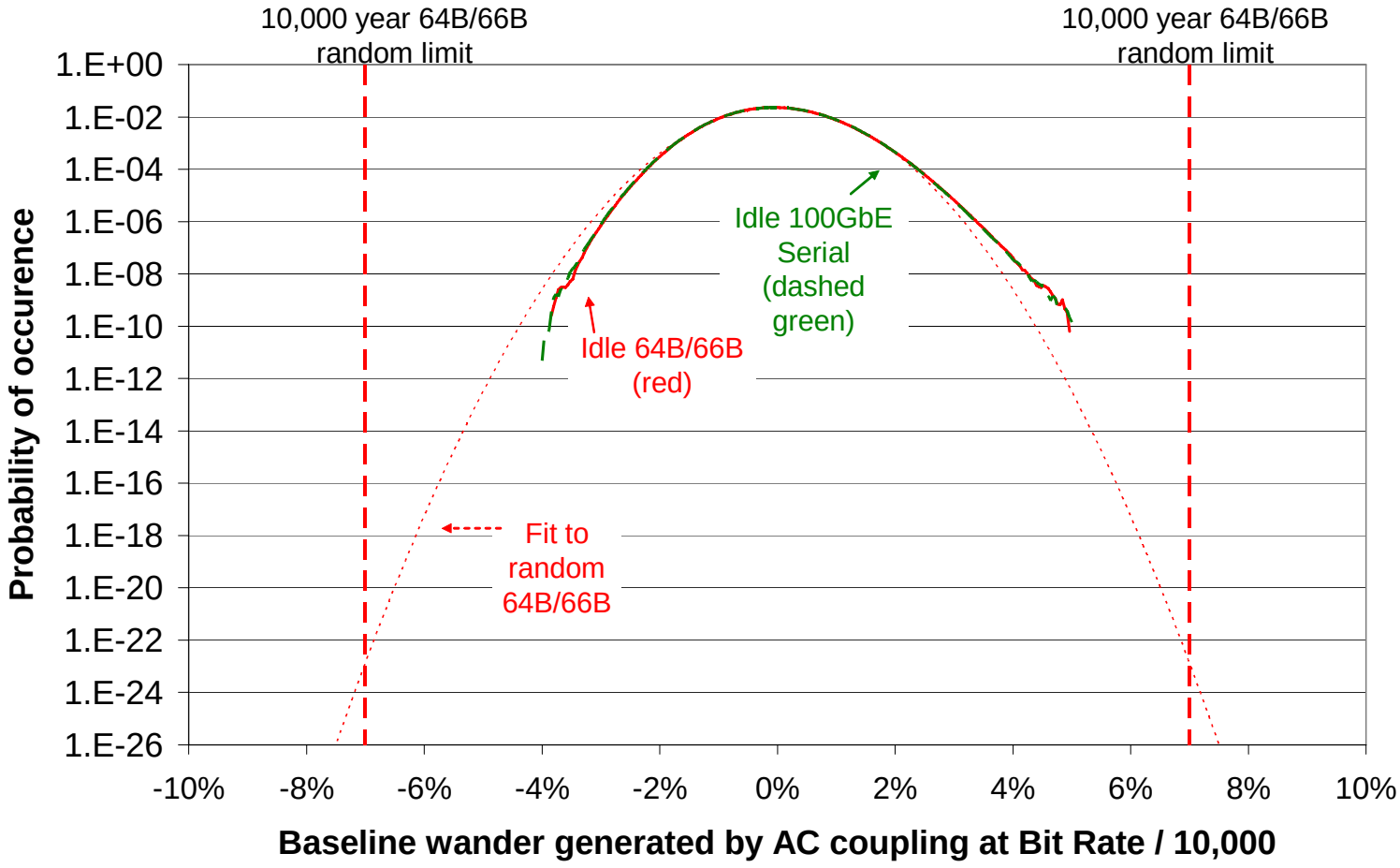
Idles

The preceding plots assumed the link is heavily loaded with traffic which, together with the scrambler, looks like random data. The other extreme from this is to look at the baseline wander and clock wander when the link is completely idle.

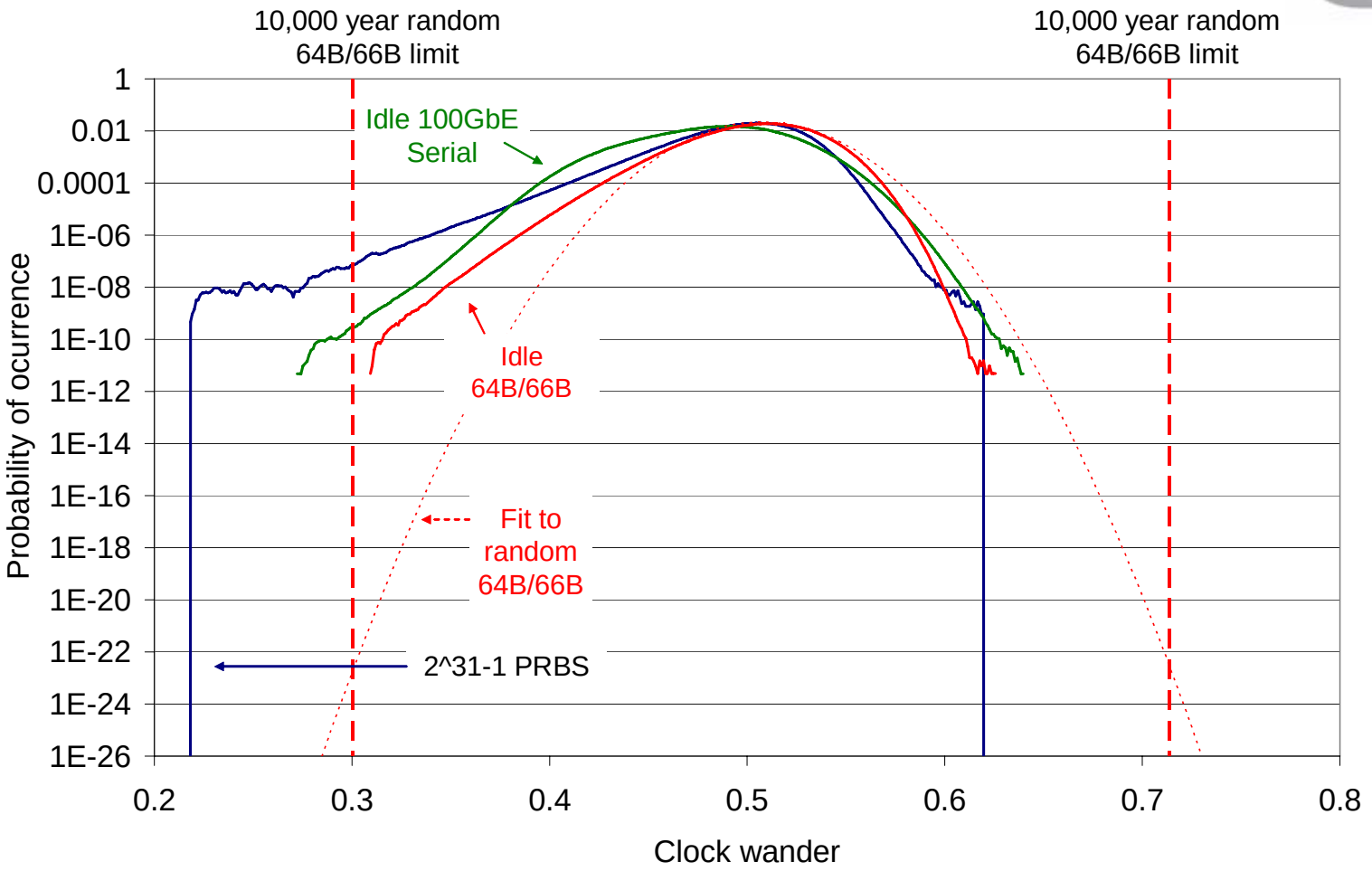
The $1 + x^{39} + x^{58}$ scrambler repeats after 3×10^{17} bits (about a month at 100 Gbit/s) so it is not feasible to simulate the entire sequence. The following plots are for 2×10^{11} bits.



Baseline wander PDF for 100G serial MLD - link idle



Clock wander PDF for 100G serial MLD - link idle



Idle results

The baseline wander PDF curve for idle 100GbE serial is still very similar to that of 64B/66B coded random data.

The clock wander PDF characteristic for idle 100GbE serial shows a similar shift towards lower clock content compared to idle 64B/66B as did the random payload 100GbE serial compared to random 64B/66B.

The trend for this curve, however, suggest that the PDF for idle 100GbE serial will remain within that for the PRBS31 test pattern.

25.8 GBd lanes for 100GbE

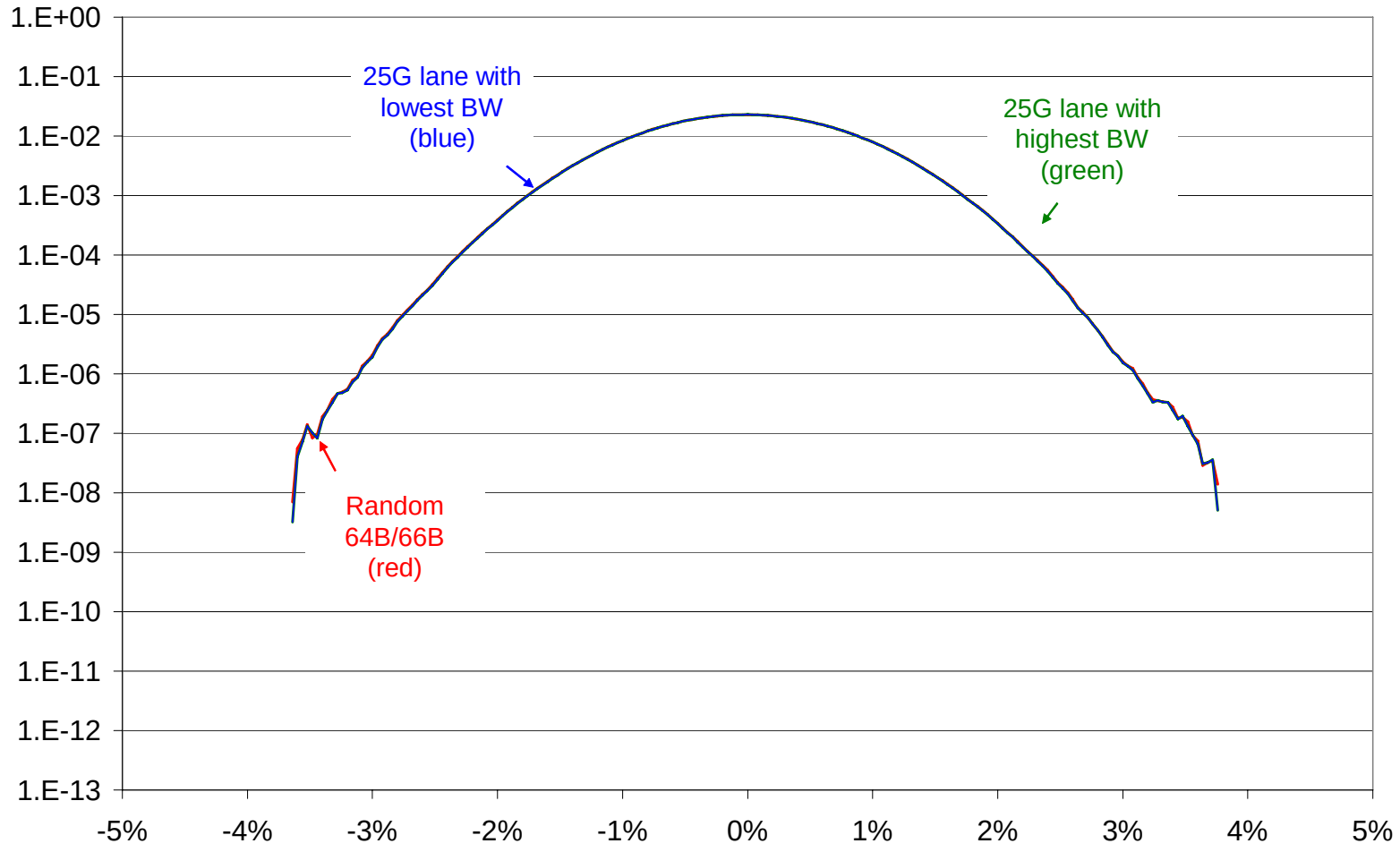
To meet the objectives for 10km and 40km over single mode fibre for 100Gb Ethernet, lanes running at 25.8 GBd are being considered. These physical lanes would each consist of 5 virtual lanes bit interleaved together.

Depending upon the implementation details of any gearbox function between say 10 electrical lanes and 4 optical lanes (and skew in the electrical lanes) the particular virtual lanes comprising each optical lane are unknown. Consequently, all possible combinations of virtual lanes were analysed to find the **worst cases** for Baseline Wander (BW) and Clock Wander (CW). These are:

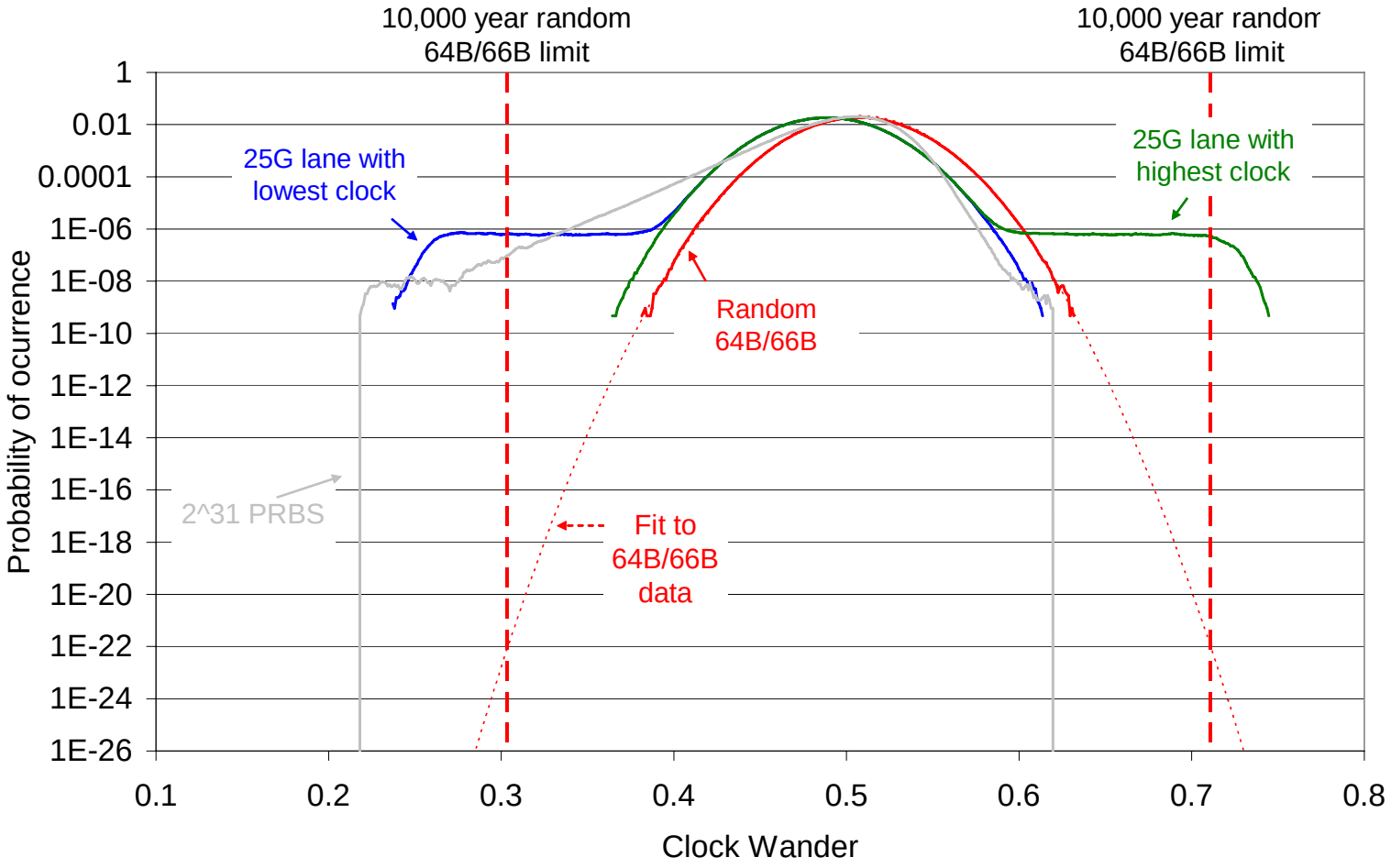
Max pos BW	VLs	1	8	15	2	4
Max neg BW	VLs	10	13	6	12	14
Max CW	VLs	5	6	9	16	18
Min CW	VLs	14	7	13	10	6

These VL combinations were then used to generate the worst case PDFs for 25.8 GBd lanes with random data.

Baseline wander PDFs for worst 25.8 GBd lanes



Clock wander PDFs for worst 25.8 GBd lanes



25 GBd results

The worst combination of VLs for a 25.8 GBd lane has a clock wander PDF that falls below that of a PRBS31 test pattern.

To address this we need to change the alignment block to a 16 or 32 bit number per lane. These versions might look as below:

Revised Alignment Block 16 bit

10	VL1	VL2	$\overline{VL1}$	$\overline{VL2}$	VL1	VL2	$\overline{VL1}$	$\overline{VL2}$

Revised Alignment Block 32 bit

10	VL1	VL2	VL3	VL4	$\overline{VL1}$	$\overline{VL2}$	$\overline{VL3}$	$\overline{VL4}$

Alignment codes

To investigate the effect of 16 bit and 32 bit alignment codes on 25.8 GBd lane properties a set of alignment codes was generated using a section of the $1 + x^{39} + x^{58}$ scrambler output with all zeros input.

Lane	VL1	VL2	Lane	VL1	VL2
0	89	CC	10	88	0E
1	C7	56	11	1D	F6
2	C8	A1	12	84	48
3	24	44	13	0E	2D
4	19	FB	14	41	36
5	0B	33	15	FC	94
6	25	1E	16	B4	19
7	ED	E0	17	AF	7A
8	75	3D	18	93	2A
9	43	E2	19	DF	85

Lane	VL1	VL2	VL3	VL4	Lane	VL1	VL2	VL3	VL4
0	C1	68	21	F4	10	FD	6C	99	DE
1	9D	71	8E	17	11	B9	91	55	B8
2	59	4B	E8	B0	12	5C	B9	B2	CD
3	4D	95	7B	10	13	1A	F8	BD	AB
4	F5	07	09	0B	14	83	C7	CA	B5
5	DD	14	C2	50	15	35	36	CD	EB
6	9A	4A	26	15	16	C4	31	4C	30
7	7B	45	66	FA	17	AD	D6	B7	35
8	A0	24	76	DF	18	5F	66	2A	6F
9	68	C9	FB	38	19	C0	F0	E5	E9

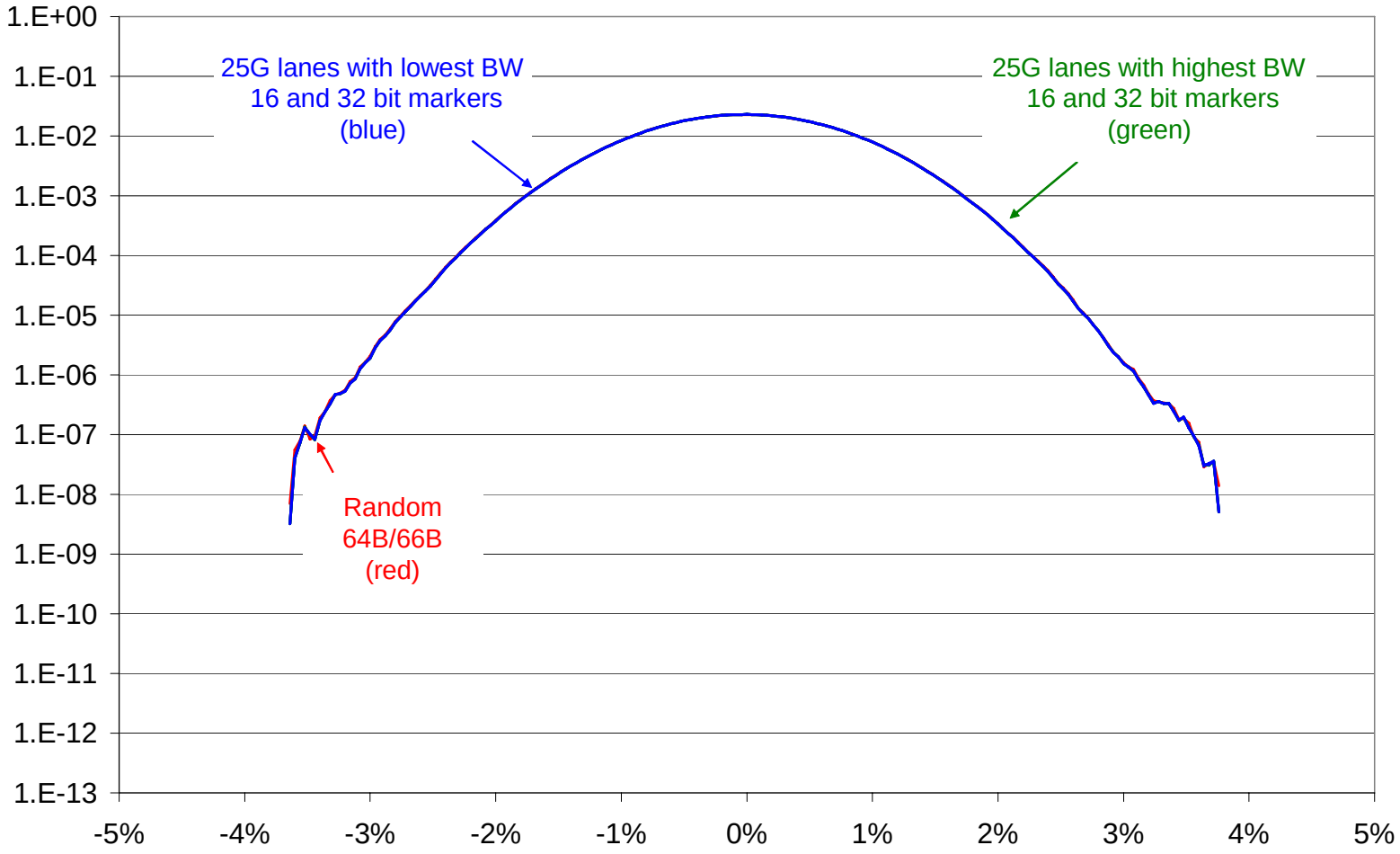
Worst case 25 GBd lanes

Using these alignment codes all possible combinations of virtual lanes were then re-analysed to find the worst cases for Baseline Wander (BW) and Clock Wander (CW). These searches also included lane delays of -1, 0 or +1 bits for each VL. The results were:

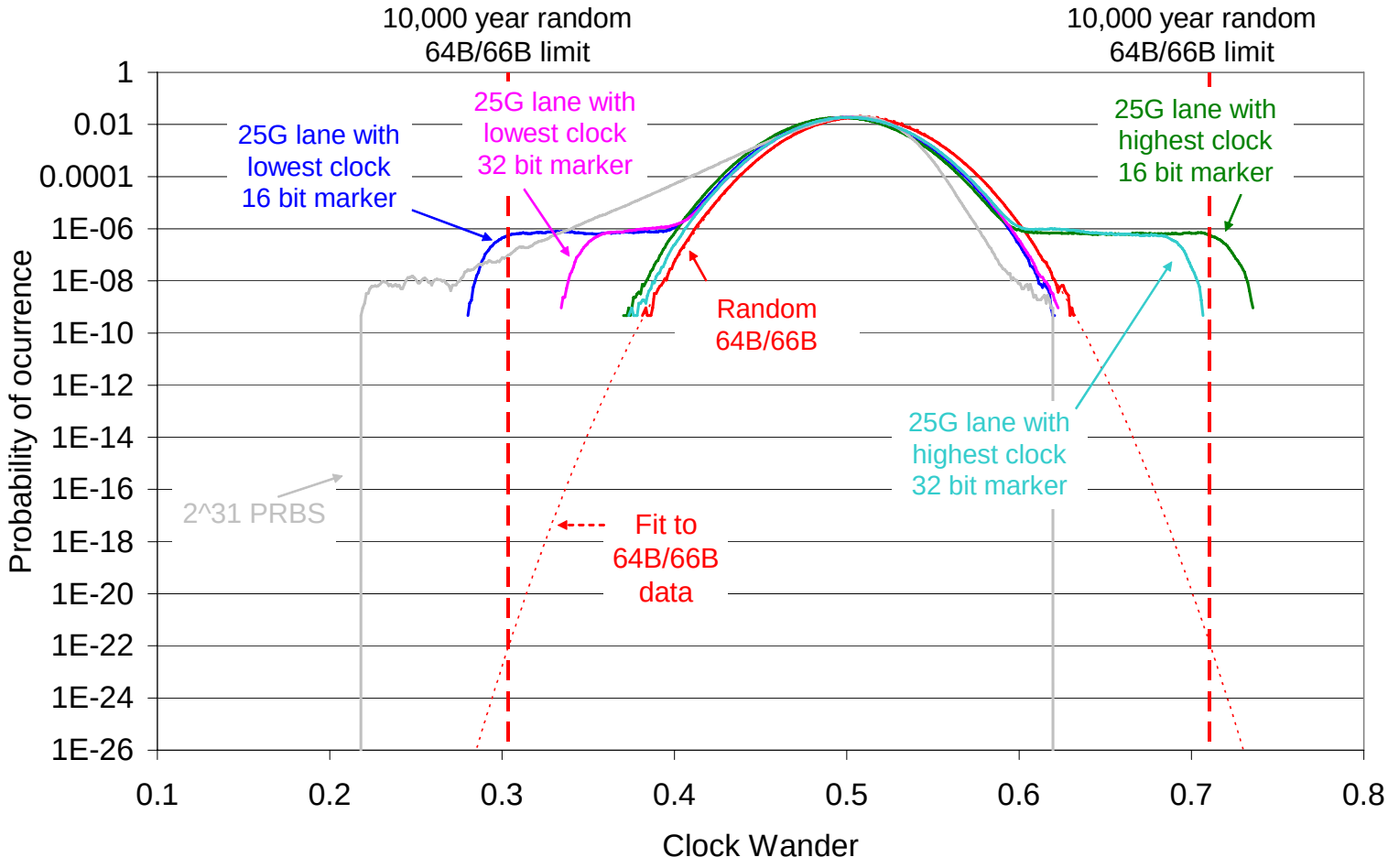
16 bit Max pos BW	VLs	10	12	3	2	9	Delays	-1	-1	-1	-1	1
16 bit Max neg BW	VLs	11	4	17	19	8	Delays	0	0	1	1	1
16 bit Max CW	VLs	16	9	6	7	0	Delays	-1	-1	-1	0	1
16 bit Min CW	VLs	3	9	14	18	12	Delays	-1	0	0	-1	1
32 bit Max pos BW	VLs	4	2	6	16	5	Delays	-1	-1	-1	0	1
32 bit Max neg BW	VLs	15	13	10	18	17	Delays	0	-1	0	0	1
32 bit Max CW	VLs	1	19	4	13	5	Delays	1	-1	0	-1	-1
32 bit Min CW	VLs	14	8	2	9	19	Delays	-1	1	0	0	-1

These VL combinations were then used to generate the worst case PDFs for 25.8 GBd lanes with random data.

Baseline wander PDFs for 25.8 GBd lanes 16 and 32 bit



Clock wander PDFs for 25.8 GBd lanes 16 and 32 bit



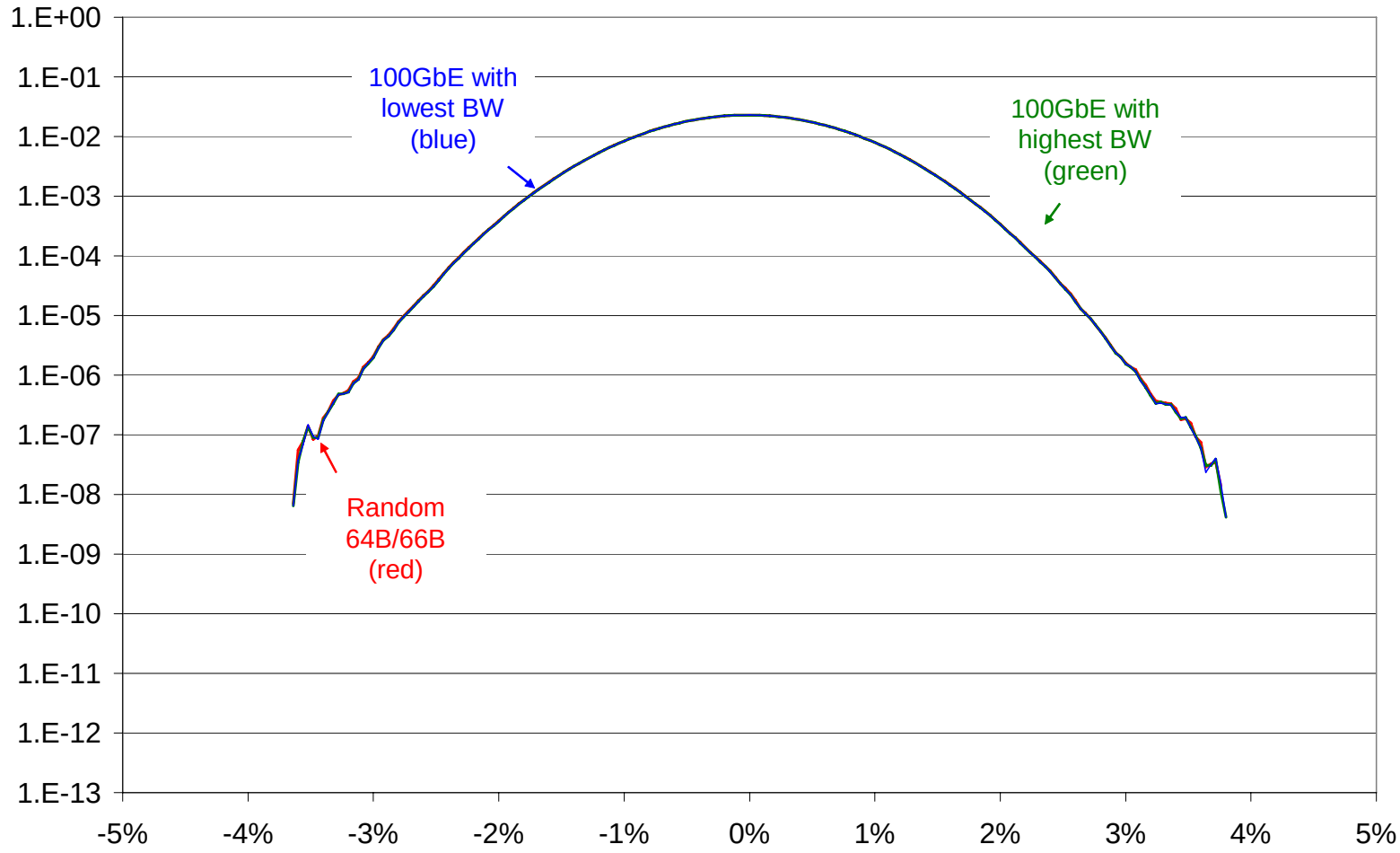
Marker length results

All four baseline wander curves (max and min for 16 bit and 32 bit markers) lie on top of the random 64B/66B curve because of the inverse following each marker.

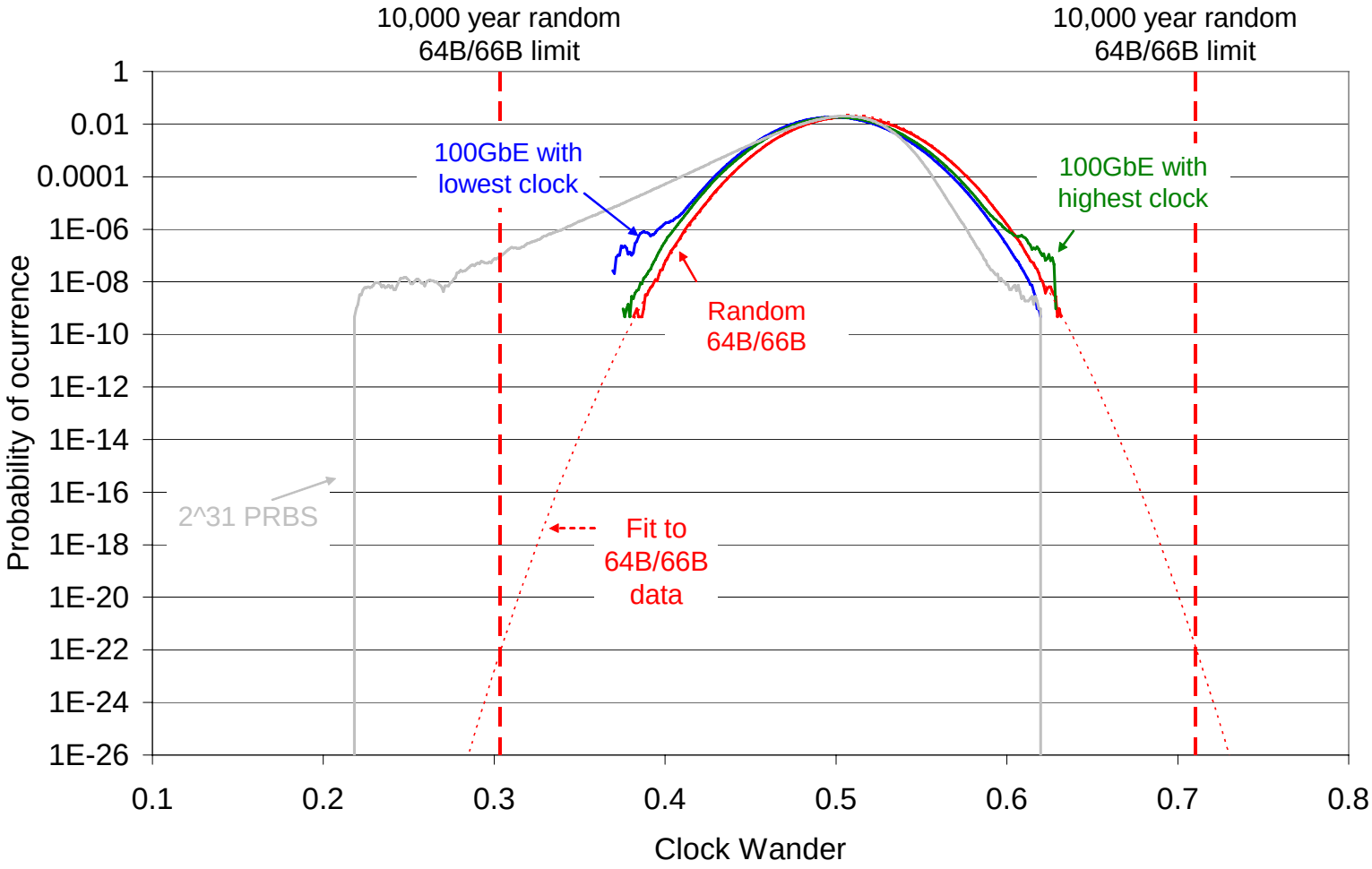
The clock wander results for 16 bit markers are better than those for the 8 bit markers. The search for the worst case, however, only included lane delays of -1, 0 or +1 bits for each VL due to restricted processing time so it is probable that a somewhat worse combination exists with larger delays. It would therefore be prudent to choose to use 32 bit lane markers to provide some margin in the clock wander behaviour for 25.8 GBd lanes for 100 GbE compared to that of a PRBS31 test pattern.

To check that the 32 bit markers also give good results with 100 GbE serial, the PDFs for this were re-calculated, this time for the worst case alignment with -1, 0 or +1 bits of skew on 10 electrical lanes.

Baseline wander PDF for 100G serial MLD - 32 bit markers



Clock wander PDF for 100G serial MLD – 32 bit markers



40 GbE serial MLD – 32 bit markers

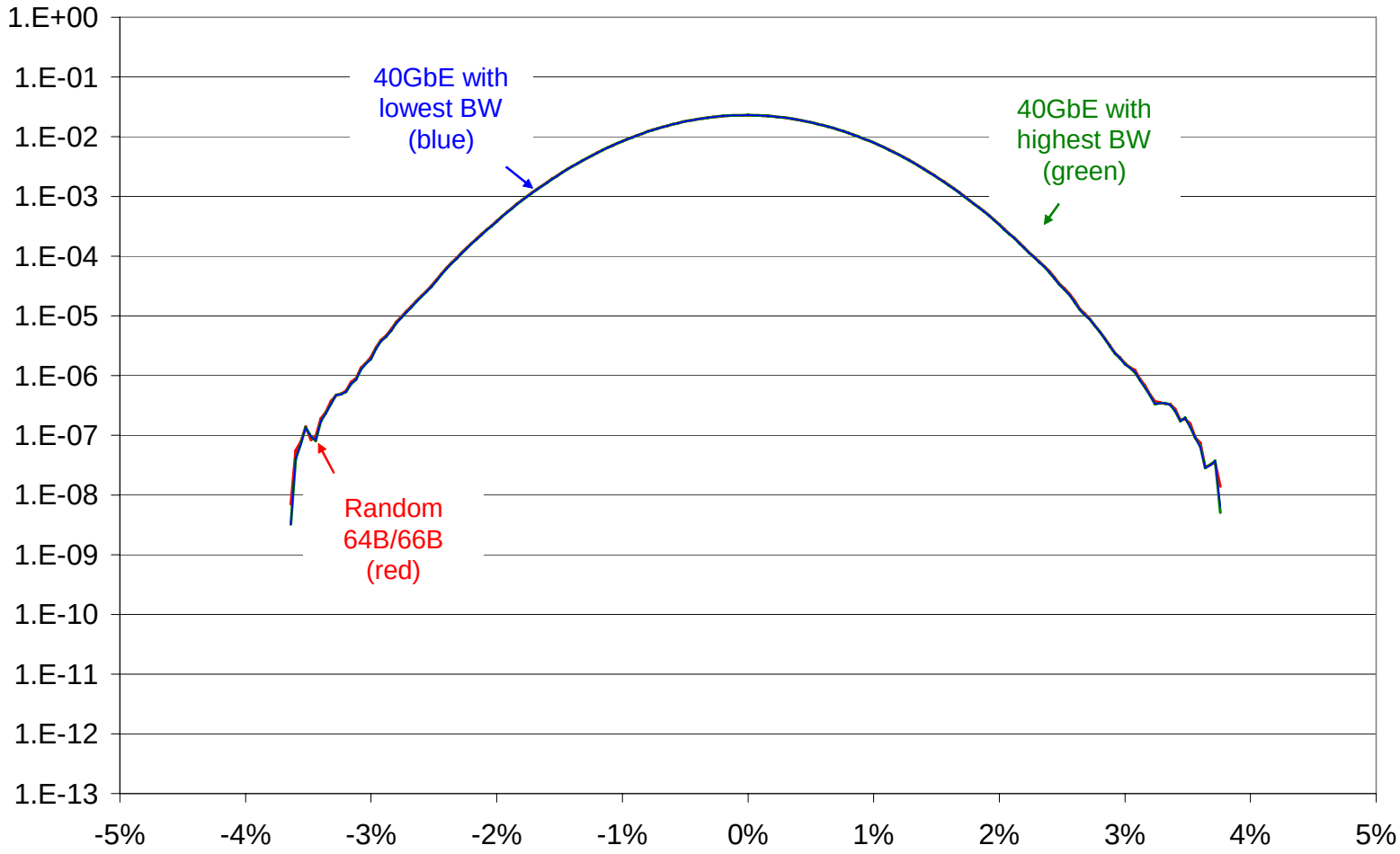
Another possible future serial PHY is 40 GbE with 4 virtual lanes with all lanes bit interleaved in to a single serial stream.

For this, all possible combinations of lanes 0 to 4 were analysed to find the worst cases for Baseline Wander (BW) and Clock Wander (CW). These searches also included lane delays of -8 to +8 bits for each VL. The results were:

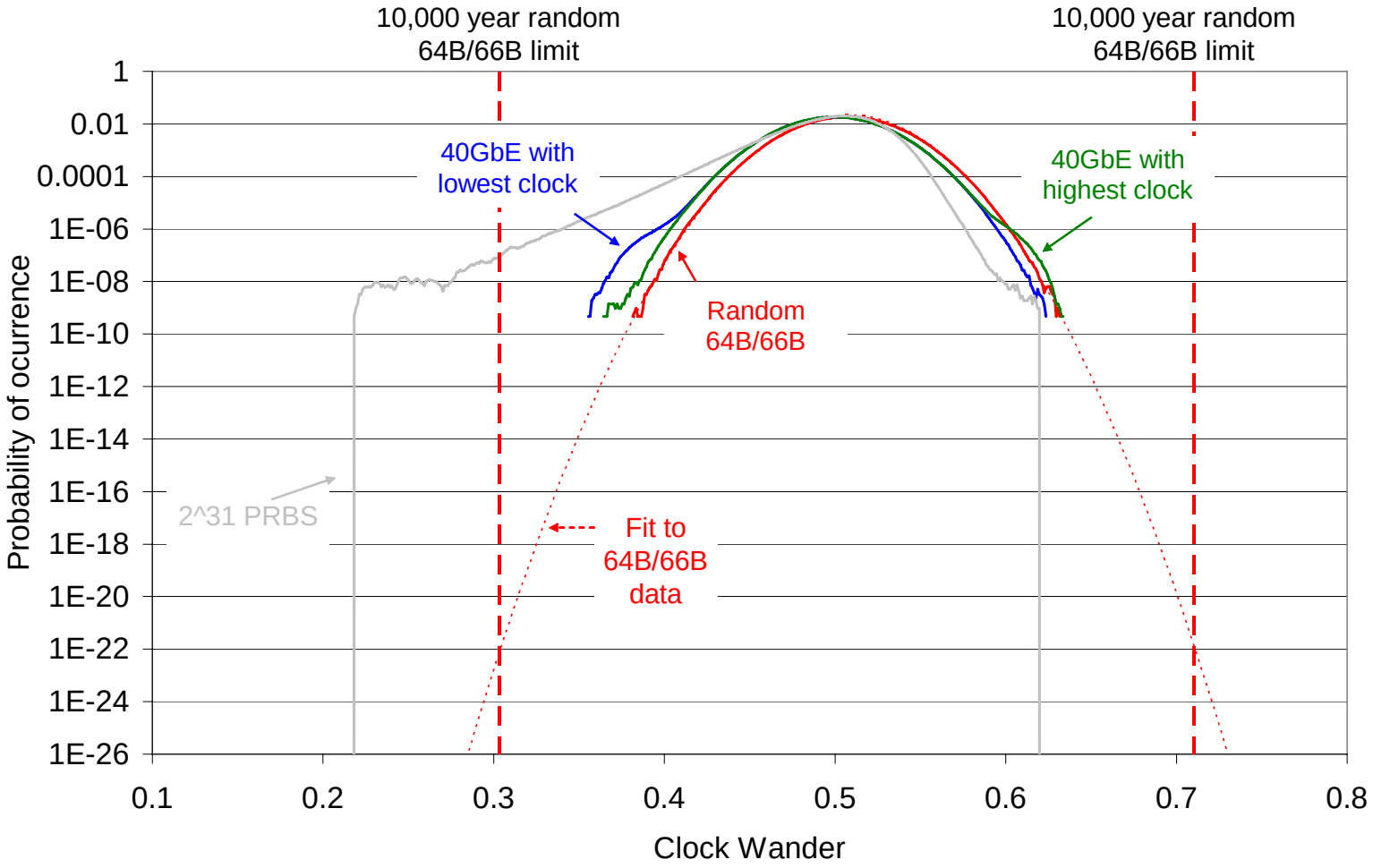
32 bit Max pos BW	VLs	0	1	2	3	Delays	5	-7	2	-6
32 bit Max neg BW	VLs	0	1	3	2	Delays	-8	3	7	5
32 bit Max CW	VLs	1	2	0	3	Delays	5	-8	-4	-2
32 bit Min CW	VLs	0	2	3	1	Delays	5	-3	8	-8

These VL combinations were then used to generate the worst case PDFs for 40 GbE serial with random data.

Baseline wander PDF for 40G serial MLD - 32 bit markers



Clock wander PDF for 40G serial MLD – 32 bit markers



Conclusions

The effect of the MLD lane alignment markers on baseline wander and clock wander has been analysed for:

- 100 GbE serial with 10 electrical lanes having -1, 0 or +1 bits of skew
- 25.8 GBd lanes of 100GbE with each VL having -1, 0 or +1 bits of skew
- 40 GbE serial with each VL having -8 to +8 bits of skew

For acceptable clock content (no worse than a PRBS31 test pattern) the lane alignment markers should contain 32 bit numbers followed by their inverse.

One possible set of 32 bit lane alignment markers is proposed based on a section of the $1 + x^{39} + x^{58}$ scrambler output with all zeros input.

Thanks!

Pete Anslow,
Nortel Networks