

in the cavity. When the cavity length is matched to the drive frequency and stabilised, the modelocked pulses have a much higher intensity than that in a mismatched cavity, and so does the polarisation trapping effect. If most energy in the cavity is trapped in one polarisation state, the laser is able to survive more polarisation changes.

With the control on, modelocked operation maintained indefinitely. To verify the stability of the control scheme, back-to-back BER measurements were carried out for the laser. At 10GHz, a BER of 10^{-9} was achieved at a received power of -18dBm (without an Er preamplifier). A long term stability test showed that no single error was detected for more than 100000s with adequate input power to the receiver, giving an error floor below 10^{-15} .

Conclusion: A very simple method has been proposed and demonstrated to stabilise modelocked erbium fibre lasers, which is based on the differential noise growth rates at different RF frequencies. Stable error free operation at 10GHz has been achieved for >100000s. Owing to this method being independent of modelocking frequency, only low frequency analogue electronic components are used, and there is no need for a computer. This technique has the potential to stabilise ultra high bit rate all optical systems, such as all optical clock recovery at 40GHz, using an erbium fibre ring laser.

© IEE 1996

25 March 1996

Electronics Letters Online No: 19960685

X. Shan and A.S. Siddiqui (Department of Electronic Systems Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom)

T. Widdowson and A.D. Ellis (BT Laboratories, Martlesham Heath, Ipswich IP5 7RE, United Kingdom)

References

- 1 WIDDOWSON, T., MALYON, D.J., SHAN, X., and WATKINSON, P.J.: 'Soliton propagation without transmission control using a phase locked erbium fibre ring laser', *Electron. Lett.*, 1994, **30**, pp. 661-663
- 2 HARVEY, G.T., and MOLLENAUER, L.F.: 'Harmonically modelocked fiber ring laser with an internal Fabry-Perot stabiliser for soliton transmission', *Opt. Lett.*, 1993, **18**, pp. 107-109
- 3 PENDER, W.A., WIDDOWSON, T., and ELLIS, A.D.: 'Error free operation of a 40Gbit/s all-optical regenerator', *Electron. Lett.*, 1996, **32**, (6), pp. 567-569
- 4 SHAN, X., CLELAND, D., and ELLIS, A.: 'Stabilising Er fibre soliton laser with pulse phase locking', *Electron. Lett.*, 1992, **28**, pp. 182-184
- 5 TAKARA, H., KAWANISHI, S., and SARUWATARI, M.: 'Stabilisation of a modelocked Er-doped fibre laser by suppressing the relaxation oscillation frequency component', *Electron. Lett.*, 1995, **31**, pp. 292-293
- 6 NAKAZAWA, M., YOSHIDA, E., and KIMURA, Y.: 'Ultrastable harmonically and regeneratively modelocked polarisation maintaining erbium fibre ring laser', *Electron. Lett.*, 1994, **30**, pp. 1603-1605
- 7 SHAN, X., ELLIS, A.D., WIDDOWSON, T., WICKENS, O.E., and SPIRIT, D.M.: 'Stabilisation of a 10GHz modelocked erbium fibre laser and its application in 4x10Gbit/s soliton transmission'. Nonlinear Guided-wave Phenomena, September 1993, Paper TuD4

Time difference simultaneous perturbation method

Y. Maeda

Indexing terms: Neural networks, Stochastic approximation, Simultaneous perturbation

The authors propose a heuristic recursive algorithm to find the minimum point of a function without using the gradient of the function. This algorithm is based on the time difference and simultaneous perturbation method. However, it does not need an additional measurement of the function to update the estimated point.

Introduction: In this Letter we consider a simple optimisation problem: finding the optimal parameter that minimises a function, while supposing that we cannot use the gradient of the function. The most basic recursive approach to the problem is the finite difference method; we add a small perturbation to a point and make measurements at the point and the perturbed point. Using a difference approximation, we can estimate the gradient of the function at the point. This estimated gradient is used to modify the point. However, if the dimension of the parameter n increases, (n large) we have to repeat this procedure n times.

The simultaneous perturbation (SP) technique overcomes some of these problems. Instead of adding perturbation to all parameters one by one, the SP technique adds perturbations to all parameters simultaneously. Then, we can obtain an estimated gradient such as the finite difference. In this procedure twice the number of measurements are required, even if n is large. This method was originally proposed by Spall [1-3], and independently devised by the author [4] and Alespector *et al.* [5]. Furthermore, some applications of this method are reported [6-8]. In this method we need twice the measurements of the function.

We propose a heuristic method based on a time difference and SP technique. Basically we use the SP idea. However, we do not make the additional measurement with a perturbed point. When we update an estimator of the optimal point, we use an estimated gradient vector and a perturbation vector simultaneously as a modifying quantity of the point. The estimated gradient of the function derives from past measurements using the SP technique. The perturbation is necessary to estimate the gradient in the next iteration.

Time difference simultaneous perturbation method: Let $\mathbf{u} \in \mathfrak{R}^n$ be a parameter vector, and $J(\mathbf{u})$ be a function to be minimised. We would like to find an optimal point $\mathbf{u}_* = \min_{\mathbf{u}} J(\mathbf{u})$. We propose the following algorithm:

$$\mathbf{u}_{t+1} = \mathbf{u}_t - \alpha \Delta \mathbf{u}_t + c \mathbf{s}_t \quad (1)$$

$$\Delta u_{t,i} = \begin{cases} \Delta u_{max} & \text{if } \frac{J(\mathbf{u}_t) - J(\mathbf{u}_{t-1})}{(cs_{t-1,i})} > \Delta u_{max} \\ -\Delta u_{max} & \text{if } \frac{J(\mathbf{u}_t) - J(\mathbf{u}_{t-1})}{(cs_{t-1,i})} < -\Delta u_{max} \\ \frac{J(\mathbf{u}_t) - J(\mathbf{u}_{t-1})}{(cs_{t-1,i})} & \text{otherwise} \end{cases} \quad (2)$$

where $\Delta u_{t,i}$ denotes the i th component of the vector $\Delta \mathbf{u}_t$. α is a positive gain coefficient. s_t is a sign vector whose components are +1 or -1 and $E(s_t) = 0$. $s_{t-1,i}$ represents the i th component of the vector s_t and $E(s_{t-1,i} s_{t-1,j}) = 0$ ($i \neq j$). c (>0) denotes the magnitude of the perturbation. By means of the term cs_t , random perturbations $+c$ or $-c$ are added to all parameters. Signs of perturbations are different in different parameters.

In eqn. 1, the vector $\Delta \mathbf{u}_t$ represents an estimated gradient vector derived from the second procedure, eqn. 2. Perturbations are added to all parameters simultaneously by the third term of eqn. 1. Since the expectation of the sign vector is zero, the estimator \mathbf{u}_t is updated only by the second term in terms of expectation.

In eqn. 2 the difference between the values of the function at time t and time $(t-1)$ is divided by the magnitude of the perturbation. This gives an estimated gradient. However, $J(\mathbf{u}_t)$ includes the effect of $\Delta \mathbf{u}_{t-1}$. When we expand $J(\mathbf{u}_t)$ at \mathbf{u}_{t-1} , there exists \mathbf{u}_m such that

$$J(\mathbf{u}_t) = J(\mathbf{u}_{t-1}) - (\alpha \Delta \mathbf{u}_{t-1} - c \mathbf{s}_{t-1})^T J'(\mathbf{u}_{t-1}) + (\alpha \Delta \mathbf{u}_{t-1} - c \mathbf{s}_{t-1})^T J''(\mathbf{u}_m) (\alpha \Delta \mathbf{u}_{t-1} - c \mathbf{s}_{t-1}) \quad (3)$$

Therefore, since $s_{t-1,i} = \pm 1$, we have

$$\begin{aligned} \Delta u_{t,i} &= \frac{J(\mathbf{u}_t) - J(\mathbf{u}_{t-1})}{c} s_{t-1,i} \\ &= \frac{(\alpha \Delta \mathbf{u}_{t-1} + c \mathbf{s}_{t-1})^T J'(\mathbf{u}_{t-1})}{c} s_{t-1,i} \\ &\quad + \frac{(\alpha \Delta \mathbf{u}_{t-1} + c \mathbf{s}_{t-1})^T J''(\mathbf{u}_m) (\alpha \Delta \mathbf{u}_{t-1} + c \mathbf{s}_{t-1})}{c} s_{t-1,i} \\ &= \frac{(\alpha \Delta u_{t-1,1} + c s_{t-1,1})}{c} s_{t-1,1} J'_1(\mathbf{u}_{t-1}) + \dots \\ &\quad + \frac{(\alpha \Delta u_{t-1,i} + c s_{t-1,i})}{c} s_{t-1,i} J'_i(\mathbf{u}_{t-1}) + \dots \\ &\quad + \frac{(\alpha \Delta u_{t-1,n} + c s_{t-1,n})}{c} s_{t-1,n} J'_n(\mathbf{u}_{t-1}) \end{aligned}$$

$$+ \frac{(\alpha \Delta \mathbf{u}_{t-1} + c \mathbf{s}_{t-1})^T \mathbf{J}''(\mathbf{u}_m)(\alpha \Delta \mathbf{u}_{t-1} + c \mathbf{s}_{t-1})}{c} s_{t-1,i} \quad (4)$$

where $J'(\mathbf{u}_{t-1}) = \partial J(\mathbf{u}_{t-1}) / \partial u_{t-1,i}$. $E(s_{t-1,i}^3) = 0$ because $s_{t-1,i}^2 = 1$ and $E(s_{t-1,i}) = 0$. Thus, the expectation of the last term of the right-hand side of eqn. 4 is zero. Moreover, since $E(s_{t-1,i} s_{t-1,j}) = 0$ ($i \neq j$) and $E(s_{t-1,i}) = 0$, taking the expectation of eqn. 4 gives

$$E(\Delta u_{t,i}) = J'_i(\mathbf{u}_{t-1}) \quad (5)$$

That is, the modifying vector Δu_t is the gradient of the function in the terms of the expectation, i.e. this procedure is a stochastic gradient method similar to the SP method. Details of this method is shown in Fig. 1.

```

begin
• Measure a value of the function  $J(\mathbf{u}_t)$ .
• Subtract the value  $J(\mathbf{u}_t)$  from the previous value of  $J(\mathbf{u}_{t-1})$ .
  (* Obtain time difference of the function *)
  for  $i := 1$  to  $n$  do
    begin
    • Multiply the time difference by  $\alpha s_{t-1,i} / c$ .
      (*  $\alpha \Delta u_{t,i}$  *)
    • Generate new sign randomly. (*  $s_{t,i}$  *)
    • Add the perturbation to  $\alpha \Delta u_{t,i}$ . (*  $\alpha \Delta u_{t,i} + c s_{t,i}$  *)
    • Update the parameter.
    end;
  Renew the iteration (*  $t = t + 1$  *)
end.

```

Fig. 1 Algorithm

Simulation: To confirm the viability of this method, we apply this to a simple multidimensional optimisation problem. The function is as follows:

$$J(\mathbf{u}) = u_1^2 + 2u_2^2 + 3u_3^2 + 4u_4^2 + 5u_5^2 + u_1u_5 + u_2u_4 + u_1u_3 \quad (6)$$

As initial values, we set up $u_{1,1} = u_{1,2} = u_{1,3} = u_{1,4} = u_{1,5} = 5.0$. The result of the time difference simultaneous perturbation method using eqns. 1 and 2 is shown in Fig. 2. The magnitude of the perturbation c is 0.01 and α is 0.001. Δu_{max} is 0.1. These are determined empirically.

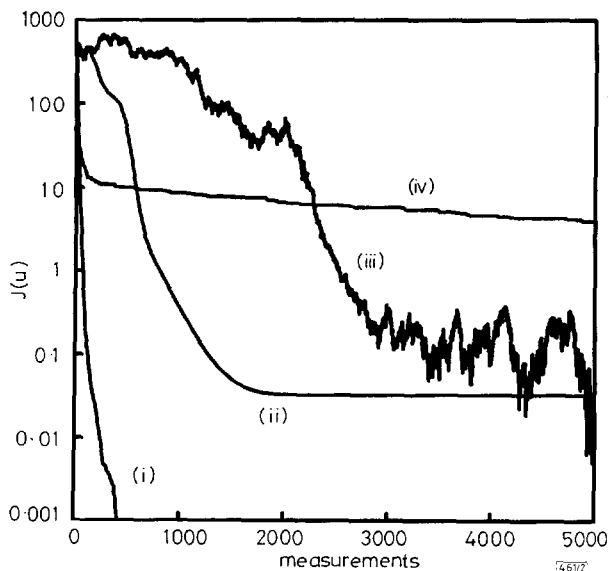


Fig. 2 Simulation result

- (i) SP
- (ii) finite difference
- (iii) time difference simultaneous perturbation
- (iv) random search

Moreover, to compare with our result, we show results using the simple random search, the finite difference technique and the SP technique. In the random search used here, random vectors in $[-5, 5]^n$ are generated using uniform distribution. The result shown in Fig. 2 is an average for 20 trials.

In the finite difference method, a difference approximation is used instead of the gradient. We fix on the constant perturbation c

(=0.1). The gain coefficient is 0.01. To guarantee the convergence of the estimator, the perturbation has to tend to zero, whether or not it is quadratic. As was expected, the estimator did not converge to zero.

Details of the SP used here are described in [4]. In this case the gain coefficient was 0.01 and the magnitude of the perturbation was 0.01. As a result, the SP method is superior to the other methods. The time difference SP method is better than the random search but inferior to the SP and the finite difference.

However, as we stated before, there are some applications where the SP or the finite difference methods are not applicable. We can provide another option. Like the other methods, this algorithm contains some adjustable coefficients. The suitable coefficients depend on an unknown objective function. However, we did find a heuristic strategy.

The role of Δu_{max} is preventing a remiss modification via an error in the estimated gradient. However, too small a value of Δu_{max} will cause a slower convergence when the estimated gradient is nearly correct or an estimator is far from an optimal point. Conversely, in the neighbourhood of the optimal point, a smaller Δu_{max} is preferable.

Eqn. 2 gives the estimated gradient using SP. Then, we must add only perturbation to the parameters and measure the value of the function. This process is eqn. 1 in my method. However, eqn. 1 includes not only the perturbation in the third term but also the modification term of the second term. This causes the error in the estimated gradient. In this sense it seems that a smaller coefficient α and/or a larger perturbation give a better estimated gradient. However, if the perturbation term is larger than the modification term in eqn. 1, modification of the parameters is not carried out properly because of the perturbation term. Good results were obtained when the magnitudes of the modification term and the perturbation term are balanced. The gradient of the function varies, depending on the position of the estimator. However, that would be a clue to determining the gain coefficient α and the perturbation c .

To guarantee the convergence of this algorithm, the perturbation c must tend to zero, because the perturbation always disturbs movements of the estimator to a proper direction in eqn. 1, and the constant perturbation does not give an exact estimation of the gradient in eqn. 2. Since the perturbation is constant in this example, it was foreseen that the estimator will remain in the neighbourhood of the optimal point.

Conclusion: A heuristic optimisation method has been proposed, based on a combination of time difference and the SP method. The ordinary finite difference needs n times measurements, and the SP method needs twice the measurements to obtain a modifying quantity for all parameters in every iteration. However, the time difference simultaneous perturbation method proposed here needs no additional measurement. We can anticipate some applications of this technique, such as neural networks. At the same time, proving the convergence of the algorithm and accelerating the convergence speed are important. For stochastic approximation we need further study.

© IEE 1996

19 February 1996

Electronics Letters Online No: 19960637

Y. Maeda (Kansai University, Faculty of Engineering, Department of Electrical Engineering, 3-3-35, Yamate-cho, Suita-chi, Osaka 564 Japan)

Y. Maeda: Currently visiting the Department of Electrical and Computer Engineering, University of California at Irvine, CA, USA

References

- 1 SPALL, J.C.: 'A stochastic approximation technique for generating maximum likelihood parameter estimates'. Proc. of the 1987 American Control Conference, 1987, pp. 1161-1167
- 2 SPALL, J.C.: 'A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting'. Proc. of the 27th IEEE Conference on Decision and Control, 1988, pp. 1544-1548
- 3 SPALL, J.C.: 'Multivariable stochastic approximation using a simultaneous perturbation gradient approximation', *IEEE Trans.*, 1992, AC-37, pp. 332-341

- 4 MAEDA, Y., HIRANO, H., and KANATA, Y.: 'A learning rule of neural networks via simultaneous perturbation and its hardware implementation', *Neural Netw.*, 1995, **8**, pp. 251-259
- 5 ALSPECTOR, J., MEIR, R., YUHAS, B., JAYAKUMAR, A., and LIPPE, D.: 'A parallel gradient descent method for learning in analog VLSI neural networks', in HANSON, S.J., COWAN, J.D., and LEE, C. (Eds.): 'Advances in neural information processing systems 5' (Morgan Kaufmann Publisher, San Mateo, CA, 1993), pp. 836-844
- 6 SPALL, J.C., and CRISTION, J.A.: 'Nonlinear adaptive control using neural networks: Estimation with a smoothed form of simultaneous perturbation gradient approximation', *Stat. Sin.*, 1994, **4**, pp. 1-27
- 7 SPALL, J.C., and CHIN, D.C.: 'A model-free approach to optimal signal light timing for system-wide traffic control'. Proc. of the 1994 IEEE Conference on Decision and Control, 1994, pp. 1868-1875
- 8 MAEDA, Y., and KANATA, Y.: 'A learning rule of neural networks for neuro-controller'. Proc. of the 1995 World Congress of Neural Networks, 1995, Vol. 2, pp. II-402-II-405

442km repeaterless transmission in a 10Gbit/s system experiment

P.B. Hansen, L. Eskildsen, S.G. Grubb, A.M. Vengsarkar, S.K. Korotky, T.A. Strasser, J.E.J. Alphonsus, J.J. Veselka, D.J. DiGiovanni, D.W. Peckham and D. Truxal

Indexing term: Optical communication

10Gbit/s unrepeated transmission is achieved over a distance of 442.5 km by using remotely pumped post and preamplifiers, low loss silica-core fibre combined with dispersion compensation, as well as artificial broadening of the spectrum to eliminate stimulated Brillouin scattering.

Repeaterless transmission at 10Gbit/s has been limited to distances up to 300km. Prechirping techniques [1] and dispersion compensating fibre [2] have been used to manage the dispersion of standard singlemode fibre resulting in reported transmission distances of 204 and 280km, respectively. A distance of 300km was reported in an experiment which included dispersion shifted fibre and a return-to-zero modulation format [3]. At 2.5 Gbit/s remotely pumped post and preamplifiers, low loss silica-core fibre combined with dispersion compensation, as well as artificial broadening of the spectrum to avoid stimulated Brillouin scattering have been instrumental in significantly increasing the maximum transmission distance of repeaterless systems [4, 5]. In this Letter we report on a repeaterless system experiment which demonstrates that these techniques are equally effective for maximising the transmission distance at a data rate of 10Gbit/s. Specifically, we have achieved a transmission distance of 442.5km with a power budget of 81.5dB.

The experimental setup is shown in Fig. 1. The transmitter consists of an externally modulated DFB laser source with a wavelength of 1556.8nm. A Ti:LiNbO₃ Mach-Zehnder modulator is employed for encoding a 2³¹-1 pseudo-random bit sequence. A Ti:LiNbO₃ phase modulator follows the intensity modulator providing artificial broadening of the spectrum to avoid stimulated Brillouin scattering. A 2.4GHz wide spectrum is generated by driving the phase modulator with three tones with frequencies of 70, 245 and 860MHz, respectively, and amplitudes corresponding to 0.9π phase shifts [6]. The power is then boosted to 20dBm by an erbium-ytterbium co-doped fibre amplifier. The transmitter terminal also includes a diode-pumped Raman laser [7] emitting 650mW at 1480nm, which is used for remotely pumping an erbium-doped fibre segment via a dedicated pump fibre. The signal output power from this remote postamplifier, which is located 74.8km into the fibre span, is 11.2dBm.

The first 125.1km of the system consists of dispersion shifted fibre with a zero-dispersion wavelength of ~1600nm to avoid modulation instability while minimising self-phase modulation. The remaining fibre employed in this experiment is low loss silica-core

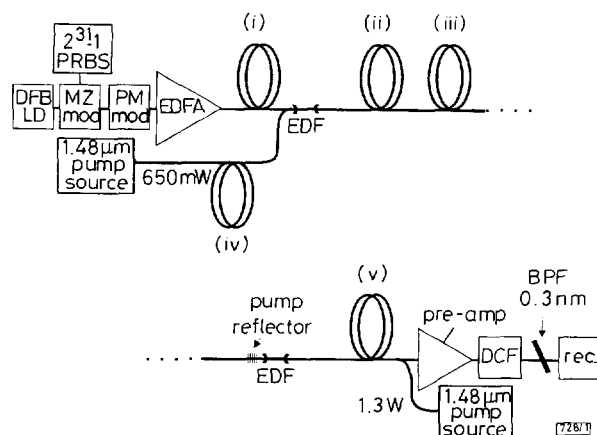


Fig. 1 Setup of 10Gbit/s repeaterless transmission experiment

Transmission distance = 442.5km; signal fibre loss = 80.6dB; losses are measured at †) 1556.8nm and ‡) 1485 nm

- (i) DSF: 74.8km (15.1dB †)
- (ii) DSF: 50.3km (10.8dB †)
- (iii) ZF: 194.4km (33.2dB †)
- (iv) ZF: 74.8km (14.7dB ‡)
- (v) ZF: 123.0km (21.5dB †)

fibre with an average loss of 0.173dB/km and a dispersion of 19.7 ps/nm/km at the signal wavelength. Table 1 summarises the length, loss and dispersion of the fibre segments which constitutes the total fibre span.

Table 1: Length, loss and dispersion of fibre segments which constitutes 442 km repeaterless transmission span DSF: dispersion shifted fibre, ZF: silica-core fibre

Span	Fibre type	Length	Loss	Dispersion
		km	dB	ps/nm
1	DSF	74.8	15.1	-240
2	DSF	50.3	10.8	-160
	ZF	194.4	33.2	3830
3	ZF	123.0	21.5	2420
		442.5	80.6	5850

An erbium doped fibre segment, which is located 123.0km from the receiver, is pumped via the transmission fibre using a 1.3W diode-pumped Raman laser. As indicated in Fig. 1, a fibre grating acts as a pump reflector for the remnant pump light from the first passage. The receiver includes a two-stage optical preamplifier, a dispersion compensating fibre module, a 0.3nm tracking Fabry-Perot filter and a *pin* detector.

The dispersion of the transmission span is compensated by dispersion compensating fibre [8], which is shown in Fig. 1 as a single DCF module. It consists of three spans of DCF each followed by an erbium-doped fibre amplifier to offset the loss. The total dispersion in the DCF module is -5190ps/nm, leaving the transmission span undercompensated by 660ps/nm.

The measured bit error rates (BER) are shown against the received power at the remote preamplifier in Fig. 2. The circles indicate the 'back-to-back' performance. The signal is injected at low power (i.e. <-20dBm) into the remote preamplifier via a 30km segment of dispersion shifted fibre in order not to change the level of Rayleigh backscattering seen by the remote preamplifier. The dispersion compensating module provides 2160ps/nm compensation for the 123km fibre span, that separates the remote preamplifier from the receiver. The measured sensitivity is -33.4 dBm for an error rate of 10⁻⁹. The squares in Fig. 2 show the measured BER for the 442.5km system as described above. A sensitivity of -33.7dBm corresponding to a negative power penalty of -0.3dB results in a power budget of 81.5dB.

In conclusion, we have demonstrated error free transmission at 10Gbit/s in an unrepeated transmission system with a fibre span of 442.5km and a transmission fibre loss of 80.6dB. The power budget for an error rate of 10⁻⁹ is 81.5dB. Technologies employed in this experiment include: (i) an SBS suppression technique,