

# Adjoint-enabled optimization of optical devices based on coupled-mode equations

Yannick Lefevre,\* Pierre Wahl, Nathalie Vermeulen, and Hugo Thienpont

*Brussels Photonics Team, Department of Applied Physics and Photonics, Vrije Universiteit Brussel, B-1050 Brussels, Belgium*

[\\*ylefevre@b-phot.org](mailto:*ylefevre@b-phot.org)

**Abstract:** In this work, we propose a method for designing optical devices described by coupled-mode equations. Following a commonly applied optimization strategy, we combine gradient-based optimization algorithms with an adjoint sensitivity analysis of the coupled-mode equations to obtain an optimization scheme that can handle a large number of design parameters. To demonstrate this adjoint-enabled optimization method, we design a silicon-on-insulator Raman wavelength converter. As structure, we consider a waveguide constructed from a series of interconnected and adiabatically-varying linear tapers, and treat the width at each interconnection point, the waveguide length, and the pump-Stokes frequency difference as independent design parameters. Optimizing with respect to these 1603 parameters results in an improvement of more than 10 dB in the conversion efficiency for a waveguide length of 6.28 cm and frequency difference 187 GHz below the Raman shift as compared to a converter designed by the conventional phase-matching design rule and operating at perfect Raman resonance. The increase in conversion efficiency is also accompanied by a more than 7 dB-improvement in the Stokes amplification. Hence, the adjoint-enabled optimization allows us to identify a more efficient method for achieving Raman conversion than conventional phase-matching. We also show that adjoint-enabled optimization significantly improves design robustness. In case of the Raman converter example, this leads to a sensitivity with respect to local variations in waveguide width that is several orders of magnitude smaller for the optimized design than for the phase-matched one.

© 2014 Optical Society of America

**OCIS codes:** (000.3860) Mathematical methods in physics; (000.4430) Numerical approximation and analysis; (190.4360) Nonlinear optics, devices; (190.5650) Nonlinear optics, Raman effect; (230.7370) Waveguides.

---

## References and links

1. J. Jensen and O. Sigmund, "Topology optimization for nano-photonics," *Laser Photon. Rev.* **5**, 308–321 (2011).
2. F. Wang, J. S. Jensen, and O. Sigmund, "Robust topology optimization of photonic crystal waveguides with tailored dispersion properties," *J. Opt. Soc. Am. B* **28**, 387–397 (2011).
3. P. Borel, A. Harpøth, L. Frandsen, M. Kristensen, P. Shi, J. Jensen, and O. Sigmund, "Topology optimization and fabrication of photonic crystal structures," *Opt. Express* **12**, 1996–2001 (2004).
4. Y. Tsuji, K. Hirayama, T. Nomura, K. Sato, and S. Nishiwaki, "Design of optical circuit devices based on topology optimization," *IEEE Photon. Technol. Lett.* **18**, 850–852 (2006).

5. J. Osgood, N. C. Panoiu, J. I. Dadap, X. Liu, X. Chen, I.-W. Hsieh, E. Dulkeith, W. M. Green, and Y. A. Vlasov, "Engineering nonlinearities in nanoscale optical systems: physics and applications in dispersion-engineered silicon nanophotonic wires," *Adv. Opt. Photon.* **1**, 162–235 (2009).
6. H. Rong, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang, and M. Paniccia, "A continuous-wave raman silicon laser," *Nature* **433**, 725–728 (2005).
7. M. A. Foster, A. C. Turner, R. Salem, M. Lipson, and A. L. Gaeta, "Broad-band continuous-wave parametric wavelength conversion in silicon nanowaveguides," *Opt. Express* **15**, 12949–12958 (2007).
8. V. Raghunathan, R. Claps, D. Dimitropoulos, and B. Jalali, "Parametric Raman wavelength conversion in scaled silicon waveguides," *J. Lightwave Technol.* **23**, 2094–2102 (2005).
9. J. B. Driscoll, N. Ophir, R. R. Grote, J. I. Dadap, N. C. Panoiu, K. Bergman, and R. M. Osgood, "Width-modulation of Si photonic wires for quasi-phase-matching of four-wave-mixing: experimental and theoretical demonstration," *Opt. Express* **20**, 9227–9242 (2012).
10. D. T. Tan, P. C. Sun, and Y. Fainman, "Monolithic nonlinear pulse compressor on a silicon chip," *Nat. Commun.* **1**, 116 (2010).
11. A. W. Snyder and J. D. Love, *Optical Waveguide Theory* (Chapman and Hall, 1983).
12. L. Jin, W. Jin, J. Ju, and Y. Wang, "Coupled local-mode theory for strongly modulated long period gratings," *J. Lightwave Technol.* **28**, 1745–1751 (2010).
13. W.-P. Huang and J. Mu, "Complex coupled-mode theory for optical waveguides," *Opt. Express* **17**, 19134–19152 (2009).
14. G. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. (Academic, 2001).
15. Q. Lin, O. J. Painter, and G. P. Agrawal, "Nonlinear optical phenomena in silicon waveguides: Modeling and applications," *Opt. Express* **15**, 16604–16644 (2007).
16. Q. Lin, J. Zhang, P. M. Fauchet, and G. P. Agrawal, "Ultrabroadband parametric generation and wavelength conversion in silicon waveguides," *Opt. Express* **14**, 4786–4799 (2006).
17. J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, 1999).
18. Y. Cao, S. Li, L. Petzold, and R. Serban, "Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution," *SIAM J. Sci. Comput.* **24**, 1076–1089 (2003).
19. R. Serban and A. C. Hindmarsh, "CVODES, the sensitivity-enabled ODE solver in SUNDIALS," in "Proceedings of the 5th International Conference on Multibody Systems, Nonlinear Dynamics and Control, Long Beach, CA" (2005).
20. P. Wahl, D. S. Ly Gagnon, C. Debaes, J. Van Erps, N. Vermeulen, D. A. B. Miller, and H. Thienpont, "B-CALM: an open-source multi-GPU-based 3D-FDTD with multi-pole dispersion for plasmonics," *Prog. Electromagn. Res.* **138**, 467–478 (2013).
21. Y. Elesin, B. Lazarov, J. Jensen, and O. Sigmund, "Design of robust and efficient photonic switches using topology optimization," *Phot. Nano. Fund. Appl.* **10**, 153–165 (2012).
22. J. S. Jensen, "Topology optimization of nonlinear optical devices," *Struct. Multidisc. Optim.* **43**, 731–743 (2011).
23. N. Vermeulen, C. Debaes, and H. Thienpont, "Coherent anti-Stokes Raman scattering in Raman lasers and Raman wavelength converters," *Laser Photon. Rev.* **4**, 656–670 (2010).
24. R. Claps, V. Raghunathan, D. Dimitropoulos, and B. Jalali, "Anti-Stokes Raman conversion in silicon waveguides," *Opt. Express* **11**, 2862–2872 (2003).
25. P. Koonath, D. R. Solli, and B. Jalali, "High efficiency CARS conversion in silicon," in "Conference on Lasers and Electro-Optics and on Quantum Electronics and Laser Science" (2008), pp. 1–2.
26. Y. Lefevre, N. Vermeulen, and H. Thienpont, "Quasi-phase-matching of four-wave-mixing-based wavelength conversion by phase-mismatch switching," *J. Lightwave Technol.* **31**, 2113–2121 (2013).
27. Y. Lefevre, N. Vermeulen, C. Debaes, and H. Thienpont, "Optimized wavelength conversion in silicon waveguides based on "off-Raman-resonance" operation: extending the phase mismatch formalism," *Opt. Express* **19**, 18810–18826 (2011).
28. R. Claps, D. Dimitropoulos, V. Raghunathan, Y. Han, and B. Jalali, "Observation of stimulated Raman amplification in silicon waveguides," *Opt. Express* **11**, 1731–1739 (2003).
29. R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.* **23**, 123–129 (1987).
30. D. Dimitropoulos, R. Jhaveri, R. Claps, J. C. S. Woo, and B. Jalali, "Lifetime of photogenerated carriers in silicon-on-insulator rib waveguides," *Appl. Phys. Lett.* **86**, 071115 (2005).
31. X. Chen, N. Panoiu, and R. Osgood, "Theory of Raman-mediated pulsed amplification in silicon-wire waveguides," *IEEE J. Quantum Electron.* **42**, 160–170 (2006).
32. E. Golovchenko, P. Mamyshev, A. Pilipetskii, and E. Dianov, "Mutual influence of the parametric effects and stimulated Raman scattering in optical fibers," *IEEE J. Quantum Electron.* **26**, 1815–1820 (1990).
33. ePIXfab, The silicon photonics website, <http://www.epixfab.eu/>.
34. T. Shoji, T. Tsuchizawa, T. Watanabe, K. Yamada, and H. Morita, "Low loss mode size converter from 0.3  $\mu\text{m}$  square Si wire waveguides to singlemode fibres," *Electron. Lett.* **38**, 1669–1670 (2002).
35. V. R. Almeida, R. R. Panepucci, and M. Lipson, "Nanotaper for compact mode conversion," *Opt. Lett.* **28**, 1302–1304 (2003).
36. D. Zografopoulos, R. Beccherelli, and E. Kriezis, "Quasi-soliton propagation in dispersion-engineered silicon

- nanowires,” *Opt. Commun.* **285**, 3306–3311 (2012).
37. O. Tsilipakos, D. C. Zografopoulos, and E. E. Kriezis, “Quasi-soliton pulse-train propagation in dispersion-managed silicon rib waveguides,” *IEEE Photon. Technol. Lett.* **25**, 724–727 (2013).
38. We employed the commercial software package MODE Solutions by Lumerical to calculate the dispersion characteristics and mode profiles of SOI waveguides.

## 1. Introduction

Optimization algorithms have proven to be a powerful tool for designing optical devices. For instance, in the field of nanophotonics, topology optimization has recently enabled the design of components with no constraints on their geometrical shapes [1]. This approach treats the material distribution as a design parameter, and employs repeated finite-element or -difference analyses and gradient-based optimization updates in combination with an adjoint approach for efficient gradient calculations. It has been successfully applied to improve a large variety of devices, including photonic crystal waveguides with tailored dispersion characteristics [2], broadband photonic-crystal waveguide bends [3], and 90° nanowaveguide bends and splitters [4].

For many other classes of optical devices there are however no such optimization schemes available. For instance, a wide variety of optical components, including nonlinear devices such as Raman amplifiers and lasers [5, 6], wavelength converters [5, 7–9], supercontinuum generators [5], pulse compressors [10], and signal regenerators [5], as well as coupled [11] and perturbed [11–13] waveguide structures are described by coupled-mode equations derived from a modal analysis [5, 11, 14]. Currently, such devices are still designed based on physical insight, or by sweeping a few design parameters. As a consequence, the full potential of these components is often not realized as the optimization remains restricted to a handful of parameters.

In this paper, we propose an optimization scheme for designing coupled-mode-based optical devices that depend on a large number of design parameters. Similar to how topology optimization combines gradient-based updates with an adjoint approach, our scheme combines gradient-based optimization algorithms with an adjoint sensitivity analysis of the coupled-mode equations describing the light propagation. In Sections 2 and 3, we describe this adjoint-enabled optimization and its mathematical aspects in detail. In Section 4, we illustrate the potential of the design method by considering a non-trivial design problem, namely the design of a silicon-on-insulator (SOI) Raman wavelength converter, in which multiple modes couple through a variety of interactions. We compare a converter designed by the conventional phase-matching design rule, with one designed through adjoint-enabled optimization. For the optimization, we consider a waveguide constructed from a series of interconnected linear tapers, and treat the waveguide width at each interconnection point, as well as the waveguide length and the pump-Stokes frequency difference, as independent design parameters. By investigating the conversion process occurring in the optimized design, we identify an alternative and more efficient method than conventional phase-matching for achieving efficient Raman conversion. We also show that adjoint-enabled optimization significantly improves the robustness with respect to the variations in the design parameters. Finally, we give our conclusions in Section 5.

## 2. Adjoint-enabled optimization of coupled-mode-based devices

A wide variety of optical devices employ waveguides to guide the flow of light. In these structures, light propagation can be described by a modal analysis, i.e., by a decomposition of the electromagnetic radiation into the harmonic modes at the considered frequencies  $\omega_1, \omega_2, \dots, \omega_N$ :

$$\mathbf{E}(\mathbf{r}, t) = \text{Re} \left[ \sum_{j=1}^N A_j(z) \mathbf{e}_j(x, y) e^{-i\omega_j t} \right]. \quad (1)$$

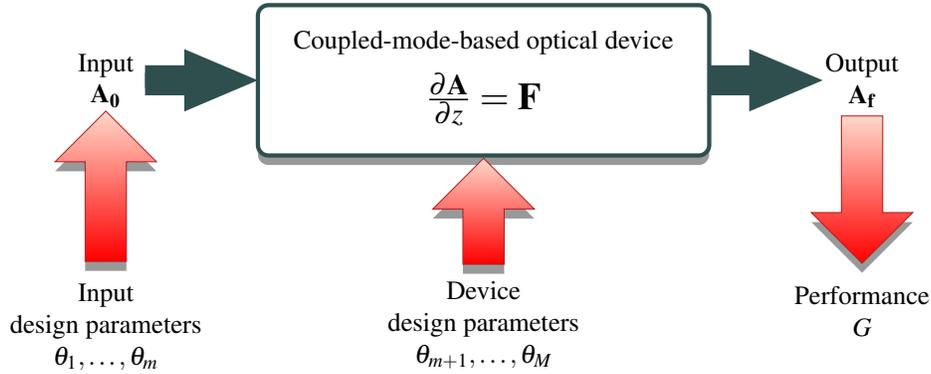


Fig. 1. A coupled-mode-based device is simulated by solving the coupled-mode equations for a given input vector  $\mathbf{A}_0$ , yielding the output vector  $\mathbf{A}_f$ . The input  $\mathbf{A}_0$  as well as the coupled-mode equations depend on a set of design parameters  $\theta_1, \dots, \theta_M$ , whereas the device's performance figure  $G$  is calculated from the output  $\mathbf{A}_f$ . Designing the device consists of determining the parameter values  $\bar{\theta}$  that optimize  $G$ .

Here the amplitudes  $A_j$  vary along  $z$ , the propagation direction of the waveguide, and the mode profiles  $\mathbf{e}_j$  capture the transverse variation as well as the polarization of each mode. The mode profiles are conventionally normalized such that  $|A_j|^2$  equals the optical power flow  $P_j$  of each mode along  $z$ . For notational simplicity, we denote by  $\mathbf{A}$  the column vector composed of the amplitudes  $A_j$ :

$$\mathbf{A} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{bmatrix}. \quad (2)$$

The light propagation is then fully specified by the evolution of the amplitude vector  $\mathbf{A}$  along  $z$ . In general, this evolution is described by Maxwell's equations, which for a wide variety of applications can be reduced to a set of coupled-mode equations, i.e., a set of coupled, first-order ordinary differential equations [11, 13–15]:

$$\frac{\partial \mathbf{A}}{\partial z} = \mathbf{F}(z, \mathbf{A}, \mathbf{A}^*, \boldsymbol{\theta}), \quad (3a)$$

$$\mathbf{A}(z_0) = \mathbf{A}_0(\boldsymbol{\theta}). \quad (3b)$$

Here  $\mathbf{A}_0$  indicates the value of the amplitude vector  $\mathbf{A}$  at the initial position  $z_0$ , and the elements of  $\mathbf{F}$  represent the coupled-mode equations of the corresponding elements of  $\mathbf{A}$ . Generally, each coupled-mode equation depends on the position  $z$ , on the full amplitude vector  $\mathbf{A}$  and its complex conjugate as denoted by the superscript  $*$ , and on  $\boldsymbol{\theta}$ , the set of design parameters  $\theta_1, \theta_2, \dots, \theta_M$  that have an impact on the light propagation. The latter include both parameters that affect the input amplitude vector  $\mathbf{A}_0$ , such as the power, phase, and frequency of the different modal amplitudes, and parameters that affect the coupled-mode equations  $\mathbf{F}$  themselves, such as structural parameters of the (local) waveguide geometry or the waveguide length. Note that we explicitly included the dependency of  $\mathbf{F}$  on both  $\mathbf{A}$  and  $\mathbf{A}^*$  to emphasize that they have to be treated as two distinct variables in the context of partial derivatives since  $\partial A_j^* / \partial A_j = 0$ .

Solving Eq. (3) for a given set of design parameters  $\boldsymbol{\theta}$  allows us to determine the corresponding output amplitude vector  $\mathbf{A}_f$  at the waveguide end position  $z_f$  [see Fig. 1]. Based on  $\mathbf{A}_f$ ,

we can determine the performance corresponding to the parameter values  $\theta$ , which is typically measured by a performance figure  $G = G(\mathbf{A}_f, \mathbf{A}_f^*)$ . Note that possibly  $G$  could also depend on other variables such as the input amplitude  $\mathbf{A}_0$  or a subset of the design parameters  $\theta$ , but we do not explicitly describe these dependencies here since they are not relevant for the current discussion. In these terms, designing a device corresponds to determining the parameter values  $\bar{\theta}$  yielding an optimal  $G$ .

For simple problems, the set of coupled equations of Eq. (3) can be solved analytically. In such cases, the parameter values  $\bar{\theta}$  are easily derived from these solutions. More complex problems, including waveguides based on materials that display a variety of nonlinear interactions, such as for instance silicon [5–8, 15, 16], and perturbed waveguides with complex non-periodic spatial variations [12] require Eq. (3) to be solved numerically. The optimal parameter values  $\bar{\theta}$  then also have to be determined numerically. Often this is done by solving Eq. (3) for a wide range of parameter values and subsequently comparing the performance of each configuration. Such a scheme is however computationally very inefficient, and becomes quickly impractical as the number of parameters  $M$  rises.

Iterative optimization algorithms allow a much smarter design strategy [17]. They generate a sequence of improving estimates for the optimal parameters based on information gathered about the performance function at the previous estimates. To do so, most algorithms start by computing the value of the performance  $G$  at each estimate. Faster gradient-based algorithms such as the steepest descent method, the conjugate gradient method, or quasi-Newton methods [17], also require to compute the gradient of the performance  $dG/d\theta_k$ , i.e., the first-order derivatives of the performance with respect to each  $\theta_k$ . Note that these derivatives are equivalent to the sensitivity of the performance with respect to each parameter  $\theta_k$ . Computing the gradient of the performance hence requires a sensitivity analysis of the set of ordinary differential equations of Eq. (3).

In case of a large number of parameters, gradient-based optimization algorithms are typically combined with an adjoint approach [1]. To compute the gradient, such methods introduce a set of additional variables, called adjoint variables, as Lagrange multipliers to efficiently calculate the gradient (sensitivity) of the system. For systems described by a set of ordinary differential equations, the adjoint sensitivity analysis is typically formulated in terms of purely real variables [18, 19], in which case the number of adjoint variables introduced equals the number of independent differential equations of the problem under study.

However, for mathematical simplicity, the variables  $\mathbf{A}$  in the coupled-mode equations of Eq. (3) are typically defined as complex variables. In this paper, we generalize the adjoint sensitivity analysis of ordinary differential equations to complex variables by counting the number of differential equations in Eq. (3) twice, since  $\mathbf{A}$  and its complex conjugate  $\mathbf{A}^*$  have to be treated as independent variables. The corresponding  $2N$  adjoint variables can be collected into two column vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$ , for which we obtain, by following a similar derivation as employed in the case of purely real variables [18, 19], the adjoint system related with Eq. (3):

$$\frac{\partial \boldsymbol{\mu}}{\partial z} = - \left( \frac{\partial \mathbf{F}}{\partial \mathbf{A}} \right)^\dagger \boldsymbol{\mu} - \left( \frac{\partial \mathbf{F}^*}{\partial \mathbf{A}} \right)^\dagger \boldsymbol{\lambda}, \quad \frac{\partial \boldsymbol{\lambda}}{\partial z} = - \left( \frac{\partial \mathbf{F}}{\partial \mathbf{A}^*} \right)^\dagger \boldsymbol{\mu} - \left( \frac{\partial \mathbf{F}^*}{\partial \mathbf{A}^*} \right)^\dagger \boldsymbol{\lambda}, \quad (4a)$$

$$\boldsymbol{\mu}(z_f) = \left( \frac{\partial G}{\partial \mathbf{A}} \right)^\dagger, \quad \boldsymbol{\lambda}(z_f) = \left( \frac{\partial G}{\partial \mathbf{A}^*} \right)^\dagger, \quad (4b)$$

where the superscript  $\dagger$  indicates the conjugate transpose, and we denote by  $\partial \mathbf{F}/\partial \mathbf{A}$  the matrix with elements  $(\partial \mathbf{F}/\partial \mathbf{A})_{ij} = \partial F_i/\partial A_j$  and by  $\partial G/\partial \mathbf{A}$  the row vector  $(\partial G/\partial \mathbf{A})_j = \partial G/\partial A_j$ . The adjoint vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$  are related with  $\mathbf{A}$  and  $\mathbf{A}^*$  respectively, and have the same size as

**A:**

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}. \quad (5)$$

The adjoint system of Eq. (4) should be solved backwards in  $z$ , i.e., from  $z_f$  to  $z_0$ . Similar to the case of purely real variables [18, 19], we can compute the total sensitivity of the performance  $G$  with respect to the parameter  $\theta_k$  from the evolutions of  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$  by the formula:

$$\frac{dG}{d\theta_k} = \frac{\partial G}{\partial \theta_k} + \boldsymbol{\mu}^\dagger(z_0) \frac{\partial \mathbf{A}_0}{\partial \theta_k} + \boldsymbol{\lambda}^\dagger(z_0) \frac{\partial \mathbf{A}_0^*}{\partial \theta_k} + 2 \int_{z_0}^{z_f} \boldsymbol{\mu}^\dagger \frac{\partial \mathbf{F}}{\partial \theta_k} + \boldsymbol{\lambda}^\dagger \frac{\partial \mathbf{F}^*}{\partial \theta_k} dz. \quad (6)$$

We point out that the adjoint method can be simplified considerably in a special class of problems. For a performance function  $G$  for which the partial derivatives in Eq. (4b) satisfy  $\partial G / \partial \mathbf{A}_f^* = (\partial G / \partial \mathbf{A}_f)^*$ , we found that the second adjoint vector equals  $\boldsymbol{\lambda} = \boldsymbol{\mu}^*$  according to Eq. (4). As a consequence, only  $N$  independent adjoint variables remain, and the adjoint system of Eq. (4) simplifies to:

$$\frac{\partial \boldsymbol{\mu}}{\partial z} = - \left( \frac{\partial \mathbf{F}}{\partial \mathbf{A}} \right)^\dagger \boldsymbol{\mu} - \left( \frac{\partial \mathbf{F}^*}{\partial \mathbf{A}} \right)^\dagger \boldsymbol{\mu}^*, \quad (7a)$$

$$\boldsymbol{\mu}(z_f) = \left( \frac{\partial G}{\partial \mathbf{A}} \right)^\dagger, \quad (7b)$$

and the adjoint sensitivity equation of Eq. (6) becomes:

$$\frac{dG}{d\theta_k} = \frac{\partial G}{\partial \theta_k} + 2 \operatorname{Re} \left[ \boldsymbol{\mu}^\dagger(z_0) \frac{\partial \mathbf{A}_0}{\partial \theta_k} \right] + 2 \int_{z_0}^{z_f} \operatorname{Re} \left[ \boldsymbol{\mu}^\dagger \frac{\partial \mathbf{F}}{\partial \theta_k} \right] dz, \quad (8)$$

The adjoint sensitivity analysis now requires only a set of  $N$  coupled differential equations to be solved. Any performance function  $G$  that is purely a function of the output powers  $P_j(z_f) = |A_j(z_f)|^2$  satisfies the condition  $\partial G / \partial \mathbf{A}_f^* = (\partial G / \partial \mathbf{A}_f)^*$ . Hence, any device for which the performance is measured only in terms of output powers can be treated by the simplified adjoint analysis of Eqs. (7)–(8), enabling a much faster gradient computation for these devices than Eqs. (4)–(6).

Similar to the approach of topology optimization, we combine the adjoint sensitivity analysis with a gradient-based iterative optimization algorithm to obtain a computationally efficient optimization tool. Concretely, the resulting scheme of *adjoint-enabled optimization* consists of five steps per iteration [see Fig. 2]: (Step 1) update the parameters  $\boldsymbol{\theta}$  based on the data obtained in the previous iterations. (Step 2) For the current parameter values  $\boldsymbol{\theta}$ , simulate the forward propagation of the amplitude vector  $\mathbf{A}$  by solving the coupled-mode equations of Eq. (3). This yields the output amplitude vector  $\mathbf{A}_f$ . (Step 3) Based on  $\mathbf{A}_f$ , calculate the current performance  $G$  and its derivatives  $\partial G / \partial \mathbf{A}_f$  and  $\partial G / \partial \mathbf{A}_f^*$ . (Step 4) Starting from these derivatives, simulate the backward propagation of the adjoint variable vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$  by solving the adjoint system of Eq. (4). This requires the knowledge of the amplitude vector  $\mathbf{A}$  across the whole waveguide length. To reduce memory usage, the evolution of  $\mathbf{A}$  can be recalculated segment-wise by employing a check-pointing algorithm [19]. (Step 5) Combine the amplitude and adjoint vectors calculated to compute the gradient of the performance function  $dG/d\boldsymbol{\theta}$  by evaluating Eq. (6) for each  $\theta_k$ . If the performance function  $G$  satisfies  $\partial G / \partial \mathbf{A}_f^* = (\partial G / \partial \mathbf{A}_f)^*$ , the adjoint system of Eq. (7) and the gradient formula of Eq. (8) can instead be employed in Step 4 and Step 5 respectively. The five steps should be repeated until  $G$  converges towards an optimum value.

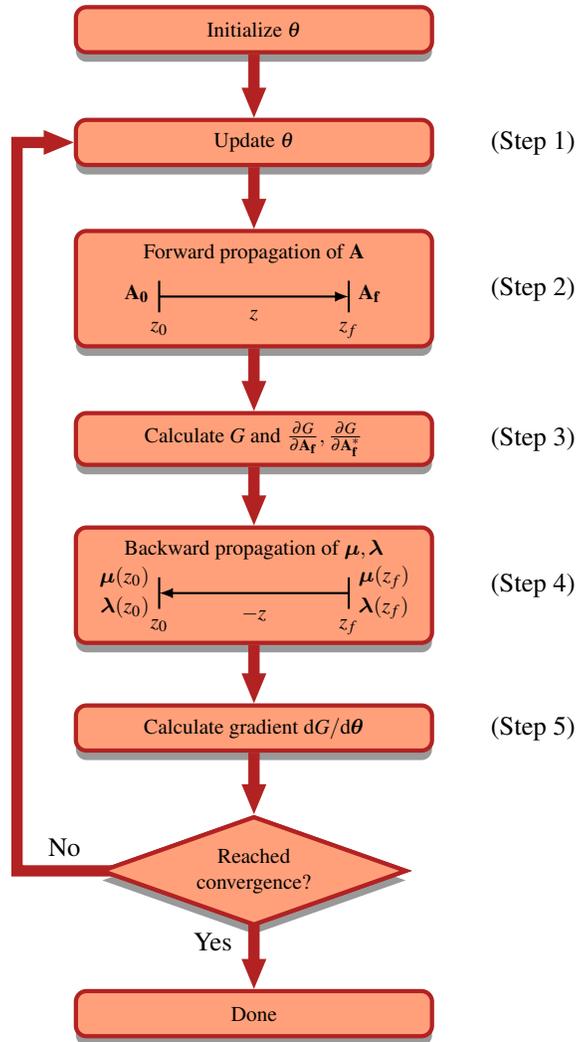


Fig. 2. Adjoint-enabled optimization optimizes the performance function  $G$  with respect to the design parameters  $\theta$  by employing an iterative optimization algorithm consisting of five steps during each iteration. Note that, if  $G$  satisfies  $\partial G / \partial \mathbf{A}_f^* = (\partial G / \partial \mathbf{A}_f)^*$ , then  $\lambda = \mu^*$  so that only one adjoint vector has to be propagated in Step 4.

### 3. Opportunities and advantages of adjoint-enabled optimization

The design method described bears similarities with the technique of topology optimization for nanophotonics [1]. The latter optimizes photonic structures by treating the material distribution as a design parameter. As in our method, this optimization is achieved by means of gradient-based optimization algorithms combined with an adjoint approach for efficient sensitivity calculations. The difference between both methods lies in how the light propagation is modeled and simulated. Topology optimization is not based on a modal analysis, but on a direct description of Maxwell's equations: the adjoint equations are directly derived from Maxwell's equations, and both sets of equations are solved with finite-element or -difference type solvers. Due to the nature of these techniques, topology optimization quickly becomes too computation-

ally intensive for devices that are large in one or more dimensions. For instance, even for highly parallel codes [20], finite-difference simulations are limited by available memory to dimensions of the order of  $(100\lambda)^3$  on supercomputers. Additionally, there exists no general framework to design nonlinear optical devices displaying intricate nonlinear interactions with topology optimization. Indeed, only nonlinear optical devices based on a Kerr-type nonlinear refractive index in a 1-D [21] and 2-D structure [22] have to our knowledge been optimized by this technique.

To overcome these limitations, our method starts from the coupled-mode equations of Eq. (3) that are approximations to Maxwell's equation. The coupled-mode equations and the derived adjoint system of Eq. (4a) or (7a) are sets of first-order ordinary differential equations that can be solved by a simple 1-D numerical integration. Hence, for optical components that are accurately described by coupled-mode equations like Eq. (3), our adjoint-enabled optimization method is much more computationally efficient than topology optimization, and thus enables the design of components that are too long to be designed by topology optimization. In addition, nonlinear optical devices based on waveguides are commonly investigated by means of coupled-mode equations [14, 15], and are thus in particular suited for our optimization method. Hence, in contrast to topology optimization, our optimization scheme can be implemented in a straightforward manner for complex nonlinear propagation equations, as we illustrate in Section 4.

These benefits are only available for geometries for which the approximations leading to the coupled-mode equations remain valid. Hence, a drawback of our adjoint-enabled optimization method is that it offers less design freedom and less geometrical flexibility than topology optimization which has no such constraints. Nevertheless, for devices that are accurately modeled by coupled-mode equations, the benefits of our method are substantial.

Generally optimization algorithms converge to local extrema and not necessarily to the sought-after global extremum. Hence, choosing a proper starting point is still essential for the design method described, and should be done with much care based on physical insight into the problem considered. By repeating the design method with multiple starting points, one can obtain a variety of locally-optimized designs that give insight into the different modes of operation improving the device's performance.

An additional advantage of employing optimization algorithms to design optical components is that the sensitivity with respect to variations in the design parameters can be greatly reduced. Indeed, at an optimum for the performance  $G$ , the gradient (sensitivity)  $dG/d\theta \approx 0$  by definition vanishes. Hence, adjoint-enabled optimization can greatly reduce the sensitivity with respect to a large number of design parameters, and allows the design of more robust devices.

It should be noted that, although we focus in this paper on waveguide devices operating in a continuous-wave regime, the described design method can also be used for designing optical components operating in a pulsed regime. Indeed, the spectral components  $A_j$  in Eq. (1) could represent the spectrum of a pulse (or of multiple pulses). Alternatively, one could also define the amplitude vector  $\mathbf{A}$  of Eq. (2) by sampling the amplitude in time rather than in frequency.

In Section 4 we illustrate the full potential of the design scheme by considering a non-trivial design problem of a component in which multiple modes couple with each other through a variety of interactions. A component that illustrates this well is the SOI-based Raman wavelength converter.

#### 4. Design of a Raman wavelength converter

Raman wavelength converters employ the third-order nonlinear Raman effect to convert light from one frequency to another [23]. convert a low-frequency Stokes wave into a high frequency anti-Stokes wave by interacting with a strong pump wave through the process of coherent anti-Stokes Raman scattering (CARS). The conversion is such that the Stokes and anti-

Stokes frequencies  $\omega_s$  and  $\omega_a$  are located symmetrically around the pump frequency  $\omega_p$ , i.e.,  $\omega_p - \omega_s = \omega_a - \omega_p$ . Due to the resonant nature of the Raman effect, the Raman interactions are only significant if the frequency detuning  $\Delta\Omega = \omega_p - \omega_s$  is close to the Raman shift  $\Delta\Omega_R$ .

Since the Raman effect is strong in silicon [15], SOI waveguides are especially suited for realizing Raman conversion in the near-infrared wavelength domain [8, 24, 25]. However, light propagation in such silicon-based waveguides is also affected by several other optical nonlinearities, namely the third-order Kerr effect and free-carrier-induced nonlinear effects. The former induces self- and cross-phase modulation, two-photon absorption (TPA), and Kerr-based four-wave mixing (FWM), and the latter the free-carrier index change (FCI) and free-carrier absorption (FCA). As a consequence, the complete coupled-mode equations modeling the pump, Stokes, and anti-Stokes propagation in SOI-based Raman converters, as derived by general methods described in literature [15], are rather complicated:

$$\begin{aligned} \frac{\partial A_p}{\partial z} = & \left( i\beta_{0,p} - \frac{\alpha_p}{2} \right) A_p + \sum_{j=p,s,a} \left( i\gamma_{K,p} (2 - \delta_{pj}) \Gamma_{jppj}^K + i\gamma_{R,p} \Gamma_{jppj}^R \right) S_j S_p |A_j|^2 A_p \\ & + \left( i\frac{\omega_p}{c} n_{f,p} - \frac{\alpha_{f,p}}{2} \right) \Gamma_p^f S_p A_p + i\gamma_{R,p} \sum_{j=s,a} \Gamma_{jppj}^R H_R (\omega_p - \omega_j) S_j S_p |A_j|^2 A_p \\ & + i \left( 2\gamma_{K,p} \Gamma_{spap}^K + \gamma_{R,p} \Gamma_{spap}^R [H_R (\omega_s - \omega_p) + H_R (\omega_a - \omega_p)] \right) S_p \sqrt{S_s S_a} A_s A_a A_p^*, \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial A_s}{\partial z} = & \left( i\beta_{0,s} - \frac{\alpha_s}{2} \right) A_s + \sum_{j=p,s,a} \left( i\gamma_{K,s} (2 - \delta_{sj}) \Gamma_{jssj}^K + i\gamma_{R,s} \Gamma_{jssj}^R \right) S_j S_s |A_j|^2 A_s \\ & + \left( i\frac{\omega_s}{c} n_{f,s} - \frac{\alpha_{f,s}}{2} \right) \Gamma_s^f S_s A_s + i\gamma_{R,s} \Gamma_{pspp}^R H_R (\omega_s - \omega_p) S_p S_s |A_p|^2 A_s \\ & + i \left( \gamma_{K,s} \Gamma_{spap}^K + \gamma_{R,s} \Gamma_{spap}^R H_R (\omega_p - \omega_a) \right) S_p \sqrt{S_s S_a} A_p^2 A_s^*, \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{\partial A_a}{\partial z} = & \left( i\beta_{0,a} - \frac{\alpha_a}{2} \right) A_a + \sum_{j=p,s,a} \left( i\gamma_{K,a} (2 - \delta_{aj}) \Gamma_{jaaj}^K + i\gamma_{R,a} \Gamma_{jaaj}^R \right) S_j S_a |A_j|^2 A_a \\ & + \left( i\frac{\omega_a}{c} n_{f,a} - \frac{\alpha_{f,a}}{2} \right) \Gamma_a^f S_a A_a + \gamma_{R,a} \Gamma_{ppaa}^R H_R (\omega_a - \omega_p) S_p S_a |A_p|^2 A_a \\ & + i \left( \gamma_{K,a} \Gamma_{spap}^K + \gamma_{R,a} \Gamma_{spap}^R H_R (\omega_p - \omega_s) \right) S_p \sqrt{S_s S_a} A_p^2 A_s^*. \end{aligned} \quad (11)$$

Here we employ a similar notation as in [26], in which  $A_j$  with  $j = p, s, a$  represent the complex amplitudes of respectively the pump, Stokes, and anti-Stokes waves. We denote the waves' propagation constants by  $\beta_j$ , and the linear losses by  $\alpha_j$ . The latter is in SOI nanowaveguides of the order  $\alpha_j = 1$  dB/cm [7]. The coefficients  $\gamma_{K,j}$  and  $\gamma_{R,j}$  describe the nonlinear Kerr and Raman effects respectively, whereas  $H_R$  is the Raman spectral response. These nonlinear parameters can be modeled by [15, 27]:

$$\gamma_{K,j} = \frac{\omega_j}{c} n_2 + i \frac{\beta_T}{2}, \quad (12)$$

$$\gamma_{R,j} = \frac{\omega_j}{\omega_{\text{ref}}} \frac{g_{R,\text{ref}} \Gamma_R}{\Omega_R}, \quad (13)$$

$$H_R(\Delta\Omega) = \frac{\Omega_R^2}{\Omega_R^2 - \Delta\Omega^2 - 2i\Gamma_R \Delta\Omega}. \quad (14)$$

Here  $\gamma_{K,j}$  is related to the Kerr coefficient  $n_2 = 6 \cdot 10^{-5}$  cm<sup>2</sup>/GW and the coefficient for TPA  $\beta_T = 0.45$  cm/GW in the near-infrared wavelength domain [16], whereas  $\gamma_{R,j}$  and  $H_R$  depend on the Raman shift  $\Omega_R = 2\pi \times 15.6$  THz, the Raman linewidth  $\Gamma_R = 2\pi \times 52.5$  GHz [16], and the Raman gain  $g_{R,\text{ref}} = 20$  cm/GW at the reference frequency  $\omega_{\text{ref}} = 2\pi c / 1542.3$  nm [28].

The coefficients  $\alpha_{f,j}$  and  $n_{f,j}$  describe the FCA and the FCI respectively. Around a reference wavelength  $\lambda_r = 1550$  nm, they are commonly related to the free-carrier density  $N_f$  by means of two empirical formulas [15, 29]:

$$\alpha_{f,j} = \left(\frac{\omega_r}{\omega_j}\right)^2 14.5 \times 10^{-18} N_f, \quad (15)$$

$$n_{f,j} = \left(\frac{\omega_r}{\omega_j}\right)^2 (-8.8 \times 10^{-4} N_f - 8.5 N_f^{0.8}) \times 10^{-18}. \quad (16)$$

In a continuous-wave regime, the density  $N_f$  generated by TPA can in turn be related to the carrier lifetime  $\tau_0$ , which typically has a value of 1 ns in SOI nanowaveguides [16, 30], and to the area of the waveguide  $A_{\text{wg}}$  by the formula [26]:

$$N_f = \sum_{i,j=p,s,a} \frac{\tau_0 (2 - \delta_{ij}) \beta_T S_i S_j \Gamma_{ijji}^K |A_i|^2 |A_j|^2}{\hbar (\omega_i + \omega_j) A_{\text{wg}}}. \quad (17)$$

For a rectangular waveguide, the waveguide area equals  $A_{\text{wg}} = hw$  with  $w$  and  $h$  the waveguide width and height respectively. The coefficients  $\Gamma_{ijkl}^K$ ,  $\Gamma_{ijkl}^R$ , and  $\Gamma_j^f$  are the Kerr, Raman, and free-carrier-induced overlap factors respectively [26]. These factors describe the impact of the various waves' mode profiles on each nonlinear effect. Finally, the factors  $S_j = c/v_{g,j} n_{S_i,j}$  represent the ratio of the modal phase velocities to the modal group velocities [26], which are also related to the ratios of the modal energy densities to the modal optical power flows [11, 26, 31].

In the terminology of Section 2, the SOI-based Raman converter is thus described by an amplitude vector  $\mathbf{A}$  consisting of three elements, namely  $A_p$ ,  $A_s$ , and  $A_a$ . Equations (9)–(11) compose the corresponding elements of the function vector  $\mathbf{F}(z, \mathbf{A}, \mathbf{A}^*, \boldsymbol{\theta})$  [see Eq. (3a)], in which the set  $\boldsymbol{\theta}$  consists of any design parameters of choice (see further). The performance of a Raman converter is measured by the conversion efficiency  $G = P_a(z_f)/P_{s,0}$ , defined as the output anti-Stokes power divided by the input Stokes power. In the following sections, we compare the performance of a conventional Raman converter designed based on the phase-matching design rule, with that of one designed by adjoint-enabled optimization.

#### 4.1. Conventional design based on phase matching

The coupled-mode equations of Eqs. (9)–(11) can be solved analytically if three assumptions are made: (1) the strong-pump assumption, which assumes that the pump power is much stronger than the Stokes and anti-Stokes powers throughout the waveguide ( $P_p \gg P_s, P_a$ ); (2) the undepleted-pump approximation, which assumes that the pump power remains approximately undepleted throughout the waveguide ( $|A_p(z)|^2 \approx P_{p,0}$ ); (3) and the assumption that the waveguide's characteristics are uniform along the waveguide length, i.e., that all parameters other than  $A_{p,s,a}$  and  $N$  in Eqs. (9)–(17) do not vary in function of  $z$ . The resulting equations can be solved analytically, and the solutions thus obtained are often employed to directly describe FWM and wavelength conversion in optical fibers [14, 32].

These solutions indicate that the conversion process strongly depends on the so-called phase-mismatch between the waves [8, 16, 27]. A small phase-mismatch results in an efficient anti-Stokes generation, whereas a large phase-mismatch suppresses the anti-Stokes generation so that amplification of the Stokes wave through stimulated Stokes Raman scattering is favoured instead [24]. Hence, according to the solutions to the simplified equations, designing a Raman wavelength converter simply corresponds to satisfying the phase-matching condition, i.e., to realizing a small phase-mismatch value. As the phase-mismatch depends on the waveguide's dispersion characteristics, such phase-matching is typically achieved by engineering the waveguide geometry [7].

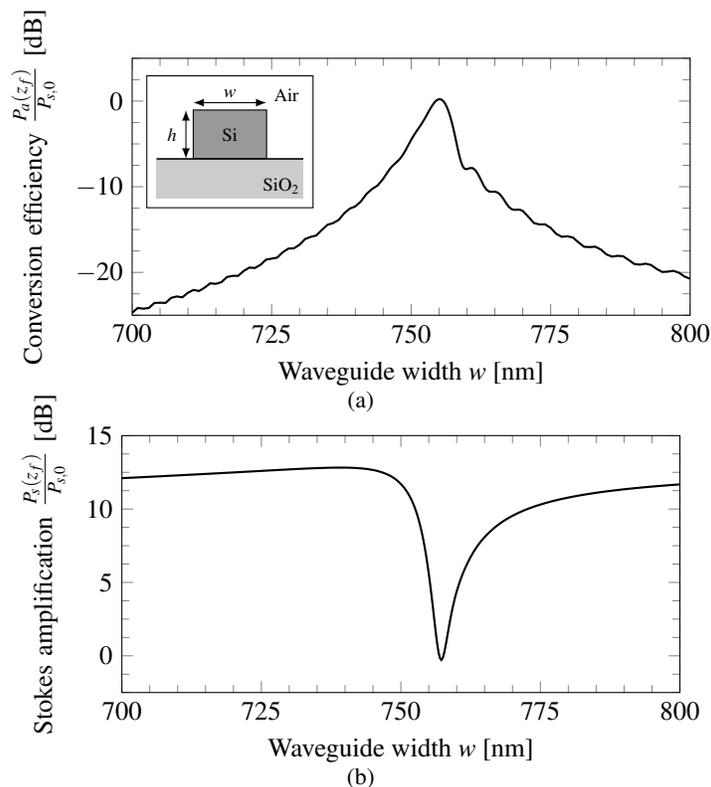


Fig. 3. (a) The conversion efficiency  $P_a/P_{s,0}$  and (b) Stokes amplification  $P_s/P_{s,0}$  of that can be achieved with a 3 cm-long, non-varying, rectangular SOI waveguide [see inset (a)] with fixed height  $h = 220$  nm depends strongly on its width  $w$ .

The phase-matching rule-of-thumb outlined above is commonly employed to design Raman wavelength converters in SOI nanowaveguides [8, 16, 24], even though the analytical solutions from which the rule was derived are actually not valid for these devices. Indeed, any near-infrared pump wave experiences severe losses in a SOI waveguide due to TPA and the associated FCA, resulting in a strong pump depletion [27], so that Eqs. (9)–(11) should be solved numerically to simulate the light propagation.

Nevertheless, the phase-matching condition remains an efficient design rule, even for these devices. To illustrate this, we consider a rectangular, air-cladded SOI waveguide [see inset Fig. 3(a)] with a non-varying waveguide geometry. We assume a waveguide height  $h$  with a fixed value of  $h = 220$  nm, as is common for silicon photonics foundries [33]. The phase-mismatch of this waveguide can be tailored by tuning the waveguide width  $w$ . For a waveguide length  $z_f = 3$  cm (we define  $z_0 = 0$ ), input pump and Stokes powers of  $P_{p,0} = 300$  mW and  $P_{s,0} = 100$   $\mu$ W, a pump wavelength of  $\lambda_p = 1550$  nm, and a pump-Stokes frequency difference at Raman resonance  $\Delta\Omega = \Omega_R$ , solving Eqs. (9)–(11) over a range of  $w$  values between 700–800 nm yields the conversion efficiency  $P_a(z_f)/P_{s,0}$  and the Stokes amplification  $P_s(z_f)/P_{s,0}$  shown in Fig. 3. As expected, a high conversion efficiency  $P_a(z_f)/P_{s,0}$  is only achieved near  $w = 755$  nm, which corresponds to a small phase-mismatch [26]. The peak in conversion efficiency is also accompanied by a dip in the Stokes amplification  $P_s(z_f)/P_{s,0}$ . For  $w$  values away from the phase-matching condition, the conversion efficiency quickly decreases,

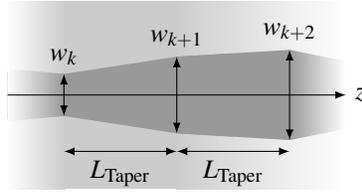


Fig. 4. The variable-width waveguide proposed consists of a series of interconnected linear tapers with equal lengths of  $L_{\text{Taper}}$ . The widths  $w_1, \dots, w_k, w_{k+1} \dots$  at each interconnection point is taken as an independent design parameter.

whereas the Stokes amplification first increases before flattening at  $w$ -values far away from phase-matching.

The phase-matching design rule allows us to easily estimate the width  $w$  of a non-varying waveguide yielding optimal conversion efficiency. In other words, it allows us to optimize the performance  $G$  with respect to but a single design parameter  $\theta_1 = w$ . Specifically, the thus optimized Raman converter design yields  $G = 0.22$  dB for  $w = 755$  nm.

#### 4.2. Design by adjoint-enabled optimization

In the previous section, only waveguides with a non-varying width along the propagation direction are considered as potential Raman converter designs. This stringent requirement is a prerequisite to obtain the analytical solutions from which the phase-matching rule is derived. However, when performing an adjoint-enabled optimization, we are not restricted by any such prerequisite. The only limitations on the waveguide design are imposed by the waveguiding functionality and by the waveguide fabrication process. If the width variations are adiabatic, then a waveguiding functionality is guaranteed. Moreover, as a planar technology, the SOI fabrication platform does allow the fabrication of SOI waveguides with arbitrary width evolutions, such as linear- [34] and parabolic-tapered [9, 35] and sinusoidally width-modulated waveguides [9]. Such SOI-based variable-width waveguides have also been proposed as a means to achieve quasi-soliton propagation [36,37] and quasi-phase-matching of FWM processes [9,26].

Due to its additional design freedom, we here also optimize a variable-width waveguide rather than a non-varying waveguide. To ensure an adiabatic width variation, we construct the waveguide as a series of interconnected linear tapers [see Fig. 4]. All tapers have an equal length  $L_{\text{Taper}}$ , so that the width evolution  $w(z)$  can be described by:

$$z_{k+1} = z_k + L_{\text{Taper}}, \quad (18a)$$

$$w(z_k \leq z \leq z_{k+1}) = w_k + \frac{w_{k+1} - w_k}{L_{\text{Taper}}} (z - z_k). \quad (18b)$$

If the taper length  $L_{\text{Taper}}$  is much larger than any variation  $w_{k+1} - w_k$  in width, then a waveguiding functionality is indeed ensured. For SOI nanowaveguides, typical width variations range from several tens to several hundreds of nanometers [9, 26, 34, 35], so that a taper length  $L_{\text{Taper}}$  of several tens of micrometers is sufficiently long [26]. For our waveguide design, we employ a taper length  $L_{\text{Taper}} = 50$   $\mu\text{m}$ .

To perform an adjoint-enabled optimization, we first have to identify the problem's design parameters. Adjoint-enabled optimization enables us to treat each width  $w_k$  in Eq. (18) as a separate design parameter. Additionally, rather than optimizing a design with a fixed length, we include the waveguide length  $z_f$  as a design parameter (we define  $z_0 = 0$ ), and impose an upper limit for this parameter. This upper limit should exceed the device length expected, and can

be specified by limiting the number of independent parameters  $w_k$  to  $M_w$  so that  $z_f \leq (M_w - 1)L_{\text{Taper}}$ . Moreover, since numerical simulations indicate that the Raman conversion efficiency can be improved by operating slightly off Raman resonance [27], we also include the frequency difference  $\Delta\Omega$  as another design parameter. The set of design parameters  $\theta$  thus consists of  $\theta_1 = \Delta\Omega$ ,  $\theta_2 = z_f$ , and  $\theta_{k+2} = w_k$  for  $k = 1, \dots, M_w$  with  $(M_w - 1)$  the maximum number of interconnected tapers allowed. Here we take  $M_w = 1601$ , resulting in a waveguide length limited by  $z_f \leq 8$  cm and a total of 1603 independent design parameters. All other parameters, including the input pump and Stokes powers, the pump wavelength, and the waveguide height, are taken as fixed values identical to those in Section 4.1, i.e.,  $P_{p,0} = 300$  mW,  $P_{s,0} = 100$   $\mu$ W,  $\lambda_p = 1550$  nm, and  $h = 220$  nm.

As iterative optimization algorithm we employ a steepest descent algorithm with a line search method based on the strong Wolfe conditions [17]. For the initial values of the design parameters, we choose a 4 cm-long waveguide operating at perfect Raman resonance and designed based on the phase-matching rule discussed in Section 4.1. This corresponds to initial values of  $\theta_1 = \Delta\Omega = \Omega_R$ ,  $\theta_2 = z_f = 4$  cm, and  $\theta_{k+2} = w_k = 755$  nm for  $k = 1, \dots, M_w$ .

During each iteration of the optimization algorithm, we execute the five steps discussed in Section 2 and depicted in Fig. 2 in the following manner: (Step 1) we update the design parameters  $\theta$  as indicated by the steepest descent algorithm. (Step 2) We solve Eqs. (9)–(11) over the current device length  $\theta_2 = z_f$ . To evaluate the  $w$ -dependent parameters in these equations, we employ the method outlined by Driscoll *et al.* [9] of fitting for each parameter a polynomial in  $w$  to a set of calculated values [38]. Evaluating the obtained polynomials at each position then allows us to directly solve the propagation equations [9, 26]. (Step 3) Based on the output anti-Stokes amplitude  $A_a(z_f)$  obtained, we update the performance  $G = |A_a(z_f)|^2/P_{s,0}$  and calculate the non-zero derivatives  $\partial G/\partial A_a(z_f) = A_a^*(z_f)/P_{s,0}$  and  $\partial G/\partial A_a^*(z_f) = A_a(z_f)/P_{s,0}$ . (Step 4) We solve the Raman converter's adjoint system in a similar fashion as the coupled-mode equations in Step 3. Since the performance derivatives satisfy  $\partial G/\partial A_a^*(z_f) = [\partial G/\partial A_a(z_f)]^*$ , we employ the simplified adjoint system described by Eq. (7). The equations for the three elements of the adjoint vector  $\mu$  are derived in a straightforward manner by applying Eq. (7) to the pump, Stokes, and anti-Stokes equations of Eqs. (9)–(11). However, as these adjoint equations are rather lengthy, we do not give them here explicitly. (Step 5) Based on the found  $\mathbf{A}$  and  $\mu$  evolutions, we calculate the gradient  $dG/d\theta_k$  for each  $\theta_k$ . First, we compute the derivative  $dG/d\Delta\Omega$  by Eq. (8). To find the function derivatives  $\partial\mathbf{F}/\partial\Delta\Omega$  of Eqs. (9)–(11) we employ the formula:

$$\frac{\partial\mathbf{F}}{\partial\Delta\Omega} = \frac{\partial\mathbf{F}}{\partial\omega_a} - \frac{\partial\mathbf{F}}{\partial\omega_s}, \quad (19)$$

which follows from the identities  $\omega_a = \omega_p + \Delta\Omega$  and  $\omega_s = \omega_p - \Delta\Omega$ . Second, we compute the derivatives  $dG/dw_k$  also by Eq. (8). The function derivatives  $\partial\mathbf{F}/\partial w_k$  are obtained by taking into account that any  $w_k$  only affects the light propagation in the tapers just before and just after the corresponding position  $z_k$ , and this according to the formulas:

$$\frac{\partial\mathbf{F}(z_k \leq z \leq z_{k+1})}{\partial w_{k+1}} = \frac{\partial\mathbf{F}}{\partial w} \frac{\partial w}{\partial w_{k+1}} = \frac{\partial\mathbf{F}}{\partial w} \frac{z - z_k}{L_{\text{Taper}}}, \quad (20)$$

$$\frac{\partial\mathbf{F}(z_k \leq z \leq z_{k+1})}{\partial w_k} = \frac{\partial\mathbf{F}}{\partial w} \frac{\partial w}{\partial w_k} = \frac{\partial\mathbf{F}}{\partial w} \frac{L_{\text{Taper}} - (z - z_k)}{L_{\text{Taper}}}. \quad (21)$$

Third, for the derivative  $\partial G/\partial z_f$ , we do not use Eq. (8), but instead employ the simple formula directly derived from  $\partial G/\partial z_f$  itself:

$$\frac{\partial G}{\partial z_f} = 2\text{Re} \left[ \frac{G}{\partial\mathbf{A}_f} \frac{\partial\mathbf{A}_f}{\partial z_f} \right] = 2\text{Re} \left[ \mu^\dagger \frac{\partial\mathbf{A}}{\partial z} \right]_{z=z_f}. \quad (22)$$

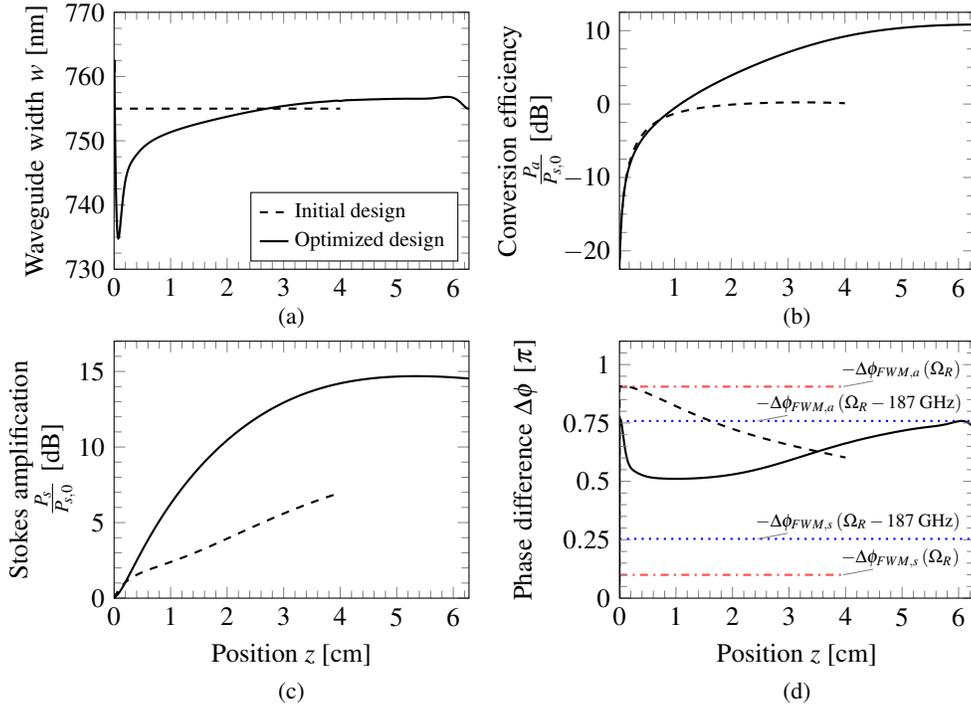


Fig. 5. Comparison between the evolutions in (a) waveguide width  $w$ , (b) conversion efficiency  $P_a/P_{s,0}$ , (c) Stokes amplification  $P_s/P_{s,0}$ , and (d) phase difference  $\Delta\phi$  for the initial Raman converter design derived from the phase-matching rule (dashed lines) and the design optimized by adjoint-enabled optimization (full lines). In (d), also the phases  $-\Delta\phi_{FWM,a}$  and  $-\Delta\phi_{FWM,s}$  of the anti-Stokes and Stokes FWM gains are shown, both for the initial frequency difference  $\Delta\Omega = \Omega_R$  (dash-dotted lines) and for the optimized one  $\Delta\Omega = \Omega_R - 187.0$  GHz (dotted lines), to indicate at which  $\Delta\phi$  values the waves experience maximal gain.

Here the last equality follows from Eq. (7b).

The resulting optimized Raman converter design is compared with the initial phase-matched design in Fig. 5. The width profile  $w$  of the optimized design varies over a range of more than 25 nm, whereas its length and frequency difference equal  $z_f = 6.28$  cm and  $\Delta\Omega = \Omega_R - 187$  GHz respectively. The variations in the width remain adiabatic, as the maximal relative change in width  $\max(|w_{k+1} - w_k|/L_{\text{Taper}}) = 4.9$  nm/50  $\mu\text{m}$  is smaller than the variation 60 nm/500  $\mu\text{m}$  of an experimentally demonstrated variable-width waveguide [9]. The optimized design's performance is  $P_a(z_f)/P_{s,0} = 10.8$  dB, corresponding to a more than 10 dB improvement with respect to the initial phase-matched design. In addition, the optimized design results in an output Stokes amplification of  $P_s(z_f)/P_{s,0} = 14.5$  dB, which is more than 7 dB higher than for the initial design. Note that the output Stokes amplification of the initial phase-matched design could also be enhanced by employing a longer waveguide, but this would be accompanied by a reduction in the conversion efficiency as the anti-Stokes power  $P_a$  experiences no longer gain but loss after 4 cm in the phase-matched converter [see Fig. 5(b)]. Hence, the optimized design does not only yield a much higher conversion efficiency than can

be achieved with the phase-matched design, but also leads to a Stokes amplification of the same level as that of a conventional Raman amplifier operating far from phase-matching [see Fig. 3]. In other words, the design combines the optimized Raman wavelength conversion with the functionality of a conventional Raman amplifier operating away from phase-matching.

To investigate the physical origins of these characteristics, we consider the evolution of the phase difference  $\Delta\phi$  along the initial and final waveguides [see Fig. 5(d)]. The phase difference, defined as  $\Delta\phi = 2\phi_p - \phi_s - \phi_a$  with  $\phi_j$  the phase of  $A_j$ , is an essential parameter in the conversion process [26, 27]. Its value determines whether the anti-Stokes and Stokes waves experience gain or loss due to the FWM processes, which consist of both CARS and Kerr-based FWM. As explained in reference [27], there is anti-Stokes (Stokes) gain as long as  $\Delta\phi$  is within a range  $\pi$  around the value  $-\Delta\phi_{FWM,a}$  ( $-\Delta\phi_{FWM,s}$ ), which is the negative of the phase of the total complex FWM anti-Stokes (Stokes) gain  $G_{FWM,a}$  ( $G_{FWM,s}$ ):

$$G_{FWM,j} = i (\gamma_{K,j} \Gamma_{spap}^K + \gamma_{R,j} \Gamma_{spap}^R H_R(\omega_j - \omega_p)) S_p \sqrt{S_s S_a} \quad (23)$$

In Fig. 5(d), we depict  $-\Delta\phi_{FWM,a}$  and  $-\Delta\phi_{FWM,s}$  both for the phase-matched design with  $\Delta\Omega = \Omega_R$  (dash-dotted lines) and for the optimized design with  $\Delta\Omega = \Omega_R - 187$  GHz (dotted lines). Conventional phase-matched operation corresponds to maintaining  $\Delta\phi$  as close as possible to  $-\Delta\phi_{FWM,a}$  so that the anti-Stokes gain is maximal throughout the waveguide. However, for the optimized waveguide design, efficient conversion is realized in a different manner entirely. Throughout the first half of the waveguide,  $\Delta\phi$  (full black line) is not maintained at the value  $-\Delta\phi_{FWM,a}$ , but rather halfway between  $-\Delta\phi_{FWM,a}$  and  $-\Delta\phi_{FWM,s}$  (blue and red dash-dotted lines respectively). As a consequence, the conversion efficiency is at the beginning of the waveguide reduced [see Fig. 5(b)] and the signal amplification increased [see Fig. 5(c)] as compared to the quantities in the initial phase-matched waveguide. However, since the anti-Stokes FWM interactions scale with  $A_s$  [27], the increased Stokes power enhances the FWM interactions further down the waveguide resulting eventually also in an increase of the conversion efficiency.

Our optimized design reveals *a posteriori* a more efficient scheme for achieving efficient Raman wavelength conversion than conventional phase-matching. Rather than maximizing the conversion locally throughout the waveguide conform the phase-matching method, the design first realizes a strong Stokes amplification. The enhanced Stokes power then enables a higher conversion efficiency towards the end of the waveguide, despite the depleted pump powers there. This scheme allows to improve the efficiency of Raman converters and even to combine conventional Raman converters and amplifiers in a single device. Additionally, it also suggests that the conversion efficiency of any phase-matched converter could potentially be improved by an initial amplification of the input signal without increasing the overall power consumption.

As discussed towards the end of Section 2, an additional advantage of a design through optimization is a reduced sensitivity with respect to the design parameters. In case of the optimized Raman converter, this translates to a reduction of the relative sensitivity  $P_a(z_f)^{-1} \partial P_a(z_f) / \partial w_k$  with respect to local variations in width by several orders of magnitude as compared to the sensitivity of the phase-matched design [see Fig. 6]. Hence, the optimized design is much more robust with respect to local fabrication errors of the waveguide width. This robustness with respect to local variations is only made possible by the adjoint-enabled optimization technique and the large number of design parameters it allows.

## 5. Conclusions

We proposed a design method for optical components that are based on coupled-mode equations. The method combines gradient-based optimization algorithms with an adjoint sensitivity

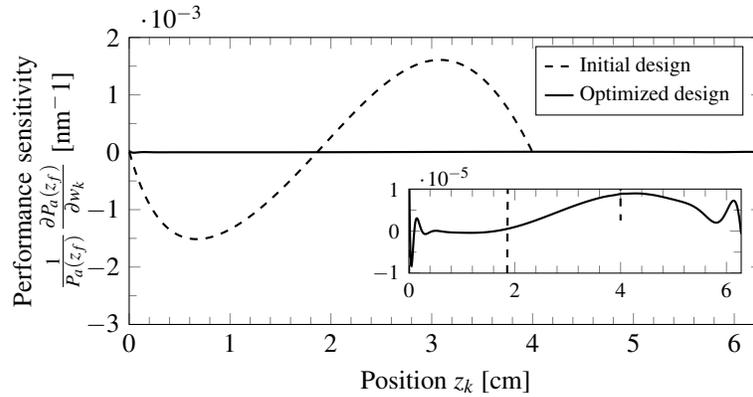


Fig. 6. The sensitivity of the Raman converter performance  $P_a(z_f)^{-1} \partial P_a(z_f) / \partial w_k$  with respect to local variations in waveguide width  $w_k$  is several orders magnitude smaller for the design optimized by adjoint-enabled optimization (full line) than for the initial design derived from the phase-matching rule (dashed line). The inset shows a close-up of the sensitivity  $P_a(z_f)^{-1} \partial P_a(z_f) / \partial w_k$  between  $-1 \cdot 10^{-5} \text{ nm}^{-1}$  and  $1 \cdot 10^{-5} \text{ nm}^{-1}$ .

analysis of the coupled-mode equations describing the light propagation to efficiently handle a large number of design parameters.

We illustrated the potential of our design method by considering the non-trivial problem of a SOI-based Raman wavelength converter that is constructed from a series of interconnected linear tapers. Optimizing with respect to 1603 design parameters, including the width at the connection points of the different tapers, the waveguide length, and the pump-Stokes frequency difference, resulted in an optimal conversion efficiency of 10.8 dB for a length of 6.28 cm and a frequency difference 187 GHz below the Raman shift. This corresponds to a more than 10 dB improvement in performance compared to a design derived from the conventional phase-matching design rule and that operates at perfect Raman resonance. Additionally, the optimized design also achieved a 14.5 dB Stokes amplification, which is more than 7 dB higher than for the phase-matched design. The adjoint-enabled optimization also allowed us to identify an alternative and more efficient method for achieving efficient Raman wavelength conversion than conventional phase-matching. By introducing a strong initial amplification of the Stokes wave, the conversion process is enhanced further down the waveguide, resulting in an overall improvement of the conversion efficiency in the optimized design. Finally, we showed that the adjoint-enabled optimization also considerably improves the design's robustness towards parameter variations. Specifically, the optimized Raman converter displays a sensitivity with respect to local variations in the waveguide width that is several orders of magnitude smaller than for the phase-matched design.

Our results show that adjoint-enabled optimization is an efficient design tool for optical components based on coupled-mode equations. The method is especially suited for non-trivial design problems that cannot be solved analytically and in which multiple modes couple through a variety of interactions. It does not only allow to improve the performance and robustness of such optical devices, but also to gain better physical insight in the mechanisms that lead to optimal performance, and even to novel classes of optical devices.

## **Acknowledgments**

This work was supported by FWO-Vlaanderen, which provides an Aspirant grant for Y. Lefevre and a Postdoctoraal Onderzoeker grant for N. Vermeulen, VUB-Methusalem, VUB-OZR, IAP-BELSPO under grant IAP P7-35, and the European Research Council (ERC-FP7/2007-2013) under grant 336940.